

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:46:18 ; Search time 902.685 Seconds  
(without alignments)  
1448.344 Million cell updates/sec

Title: US-10-716-005-1  
Perfect score: 23  
Sequence: 1 tgaagaggcagtagaagcttag 23

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues  
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : GenEmbl.\*  
1: gb.ba.\*  
2: gb.in.\*  
3: gb.env.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pr.\*  
9: gb.ro.\*  
10: gb.sts.\*  
11: gb.sy.\*  
12: gb.un.\*  
13: gb.vi.\*  
14: gb.htg.\*  
15: gb.pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	6	CS109170 Sequence
2	23	100.0	1731	6	C647508 Sequence
3	23	100.0	1734	6	AX608629 Sequence
4	23	100.0	21781	1	AE014229 Streptoco
5	23	100.0	29072	6	AX602188 Sequence
6	23	100.0	167050	1	SAG766847
7	23	100.0	349980	6	C655069 Sequence
8	23	100.0	349980	6	AX954529 Sequence
9	21	91.3	185291	14	AC123147
10	21	91.3	231447	14	AC123147
11	20.4	88.7	306616	14	AC095375
12	19.8	86.1	1378	6	AR353664 Sequence
13	19.8	86.1	1378	6	AX817901 Sequence
14	19.8	86.1	1379	6	AR353657 Sequence
15	19.8	86.1	1379	6	AR353663 Sequence
16	19.8	86.1	1379	6	AR353666 Sequence
17	19.8	86.1	1379	6	AX817894 Sequence
18	19.8	86.1	1379	6	AX817900 Sequence

19	19.8	86.1	1379	6	AX817903	AX817903 Sequence
20	19.8	86.1	1384	6	AR353658	AR353658 Sequence
21	19.8	86.1	1384	6	AR353660	AR353660 Sequence
22	19.8	86.1	1384	6	AX817895	AX817895 Sequence
23	19.8	86.1	1384	6	AX817897	AX817897 Sequence
24	19.8	86.1	1385	6	AR353659	AR353659 Sequence
25	19.8	86.1	1385	6	AX817896	AX817896 Sequence
26	19.8	86.1	1390	6	AR353661	AR353661 Sequence
27	19.8	86.1	1390	6	AX817898	AX817898 Sequence
28	19.8	86.1	1393	6	AR353665	AR353665 Sequence
29	19.8	86.1	1393	6	AX817902	AX817902 Sequence
30	19.8	86.1	1447	6	AR353667	AR353667 Sequence
31	19.8	86.1	1447	6	AX817904	AX817904 Sequence
32	19.8	86.1	1731	6	C647510	C647510 Sequence
33	19.8	86.1	1803	6	AR353656	AR353656 Sequence
34	19.8	86.1	1803	6	AX817892	AX817892 Sequence
35	19.8	86.1	10430	1	AE006575	AE006575 Streptoco
36	19.8	86.1	11992	1	AE010058	AE010058 Streptoco
37	19.8	86.1	51454	1	AE014157	AE014157 Streptoco
38	19.8	86.1	110000	1	BA000034_07	Continuation (8 of
39	19.8	86.1	110000	1	BA000034_08	Continuation (9 of
40	19.8	86.1	110000	1	CP000003_10	Continuation (11 o
41	19.8	86.1	110000	1	CP000017_10	Continuation (11 o
42	19.8	86.1	110000	1	CP000056_11	Continuation (12 o
43	19.8	86.1	140072	14	AC133332	AC133332 Felis cat
44	19.8	86.1	164494	9	AC154824	AC154824 Mus muscu
45	19.8	86.1	270432	14	CT010447	CT010447 Mus muscu
46	19.8	86.1	270432	14	CT010447	CT010447 Mus muscu
47	19.8	86.1	319008	14	AC110439	AC110439 Rattus no
48	18.8	81.7	60036	14	AC009610	AC009610 Homo sapi
49	18.8	81.7	132003	8	AC138779	AC138779 Homo sapi
50	18.8	81.7	164084	14	AC026580	AC026580 Homo sapi
51	18.8	81.7	181895	14	CT025542	CT025542 Mus muscu
52	18.8	81.7	191119	14	AC118922	AC118922 Rattus no
53	18.8	81.7	200562	9	AC117689	AC117689 Mus muscu
54	18.8	81.7	239438	14	AC107536	AC107536 Rattus no
55	18.8	81.7	239808	14	AC128575	AC128575 Rattus no
56	18.8	81.7	243280	14	AC128356	AC128356 Rattus no
57	18.4	80.0	3310	8	HSL81824	L81824 Homo sapien
58	18.4	80.0	60634	14	AC009010	AC009010 Homo sapi
59	18.4	80.0	60791	8	AC004749	AC004749 Homo sapi
60	18.4	80.0	79041	8	AC009016	AC009016 Homo sapi

## ALIGNMENTS

RESULT 1	CS109170	CS109170	Sequence 1 from Patent EPI541697.	23 bp	DNA	linear	PAT 22-JUN-2005
LOCUS	CS109170	CS109170	GI:68148019				
DEFINITION	CS109170	CS109170	GI:68148019				
ACCESSION	CS109170	CS109170	GI:68148019				
VERSION	CS109170.1	CS109170.1	GI:68148019				
KEYWORDS	synthetic construct	synthetic construct	other sequences; artificial sequences.				
SOURCE	synthetic construct	synthetic construct	other sequences; artificial sequences.				
REFERENCE	1	Uhl,J.R., Cockerill,F.R., Aichinger,C. and Reiser,A.	Detection of group B streptococcus				
AUTHORS	Uhl,J.R., Cockerill,F.R., Aichinger,C. and Reiser,A.	Patent: Bp 1541697-A 1 15-JUN-2005;	Mayo Foundation for Medical Education and Research (US)				
JOURNAL	Patent: Bp 1541697-A 1 15-JUN-2005;	Mayo Foundation for Medical Education and Research (US)	Location/Qualifiers				
FEATURES	Location/Qualifiers	Location/Qualifiers	Location/Qualifiers				
source	1..23	1..23	1..23				
ORIGIN	/organism="synthetic construct"	/organism="synthetic construct"	/mol_type="unassigned DNA"				
	/db_xref="taxon:32630"	/db_xref="taxon:32630"	/note="Oligonucleotide"				

Query Match 100.0%; Score 23; DB 6; Length 23;  
Best Local Similarity 100.0%; Pred. No. 4.7;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 1 TGAGAGGCAGTAGAAGCTTAG 23

RESULT 2
CO647508
LOCUS 1731 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 4465 from Patent WO0234771.
ACCESSION CO647508
VERSION CO647508
KEYWORDS GI:41683480
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
REFERENCE 1
AUTHORS Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 4465 02-MAY-2002;
Chiron S.p.A. (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1. .1731
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 23; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No 3.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 177 TGAGAGGCAGTAGAAGCTTAG 199

RESULT 3
AX608629
LOCUS 1734 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 6558 from Patent WO02092818.
ACCESSION AX608629
VERSION AX608629
KEYWORDS GI:28404206
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
REFERENCE 1
AUTHORS Glaser,P., Rusniok,C., Chevallier,F., Frangeul,L., Lalioui,L.,
Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
Kunst,F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 6558 21-NOV-2002;
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
1. .1734
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 23; DB 6; Length 1734;
Best Local Similarity 100.0%; Pred. No 3.2;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 177 TGAGAGGCAGTAGAAGCTTAG 199

RESULT 4
AE014229
LOCUS 21781 bp DNA linear BCT 23-SEP-2002
DEFINITION Streptococcus agalactiae 2603V/R section 39 of 100 of the complete
genome.
ACCESSION AE014229
VERSION AE014229
KEYWORDS GI:22533831
SOURCE Streptococcus agalactiae 2603V/R
ORGANISM Streptococcus agalactiae 2603V/R
REFERENCE 1
AUTHORS Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Unayam,L.A.,
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
Mariani,M., Vegni,F., Malone,D., Rinaudo,D., Rappuoli,R.,
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
TITLE Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
PUBMED 12200547
REFERENCE 2
AUTHORS Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Unayam,L.A.,
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
Mariani,M., Vegni,F., Malone,D., Rinaudo,D., Rappuoli,R.,
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. .21781
/organism="Streptococcus agalactiae 2603V/R"
/mol_type="genomic DNA"
/strain="2603V/R"
/serotype="V"
/db_xref="taxon:208435"
103..348
/gene="SAG0813"
103..348
/gene="SAG0813"
/feature="identified by Glimmer2; putative"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAW99700.1"
/db_xref="GI:22533832"
/translation="MNNQINHCFSIMVLGVVTQMKKEFTITGRMVAVRVMETLVCLCMV
SNLVDLMDIIBINEIVFCVLVASTAGLEFCLEKERE"
387..674
/gene="SAG0814"
387..674
/gene="SAG0814"
/feature="identified by Glimmer2; putative"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAW99701.1"
/db_xref="GI:22533833"
/translation="MTYIGVPKEQKALSRGKLEKINGGKTLMSQIDLVGSCFM
IYLEDGSQTHPYEILKVDWDVQIRNRLKRSKEEVRIVRIHDIKELRPVNK"
719..934
/gene

```





VERSION AX602188.1 GI:28402057  
KEYWORDS  
SOURCE Streptococcus agalactiae  
ORGANISM Streptococcus agalactiae  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1  
AUTHORS Glaeser, P., Rueniok, C., Chevalier, F., Frangeul, L., Lalioui, L., and Zouine, M., Couve, E., Buchrieser, C., Poyart, C., Trieu-Cuot, P., and Kunst, F.  
TITLE Streptococcus agalactiae genome sequence, use for developing vaccines, diagnostic tools, and for identifying therapeutic targets  
JOURNAL Patent: WO 02092818-A 117 21-NOV-2002;  
INSTITUT PASTEUR (FR) : CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)  
FEATURES  
source Location/Qualifiers  
1. 29072  
/organism="Streptococcus agalactiae"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1311"  
ORIGIN  
Query Match 100.0%; Score 23; DB 6; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 2.6;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGAGGCAGTAGAAGCTTAG 23  
|||||  
DB 12072 TGAGAGGCAGTAGAAGCTTAG 12094  
|||||  
RESULT 6  
SAG766847  
LOCUS SAG766847 167050 bp DNA linear BCT 16-APR-2005  
DEFINITION Streptococcus agalactiae NEM316 complete genome, segment 5.  
ACCESSION AL766847 AL732656  
VERSION AL766847.1 GI:23095200  
KEYWORDS  
SOURCE Streptococcus agalactiae NEM316  
ORGANISM Streptococcus agalactiae NEM316  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE 1  
AUTHORS Glaeser, P., Rueniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Msadek, T., Zouine, M., Couve, E., Lalioui, L., Poyart, C., Trieu-Cuot, P., and Kunst, F.  
TITLE Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease  
JOURNAL Mol. Microbiol. 45 (6), 1499-1513 (2002)  
PUBMED 12354221  
REFERENCE 2  
AUTHORS Glaeser, P., Rueniok, C. and Frangeul, L.  
TITLE Direct Submission  
JOURNAL Submitted (31-MAY-2002) Glaeser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86  
FEATURES  
source Location/Qualifiers  
1. 167050  
/organism="Streptococcus agalactiae NEM316"  
/mol\_type="genomic DNA"  
/strain="NEM316"  
/serovar="serovar III"  
/db\_xref="taxon:211110"  
join(288..293,303..1412)  
/gene="gbs0767"  
288..293  
/gene="gbs0767"  
303..1412  
/gene="gbs0767"  
/note="Similar to riboflavin specific deaminase (diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino)uracil

reductase)"  
/codon\_start=1  
/transl\_table=11  
/product="Unknown"  
/protein\_id="CAD46411.1"  
/db\_xref="GI:23095201"  
/db\_xref="GOA:Q8B660"  
/db\_xref="InterPro:IPR002125"  
/db\_xref="InterPro:IPR002734"  
/db\_xref="InterPro:IPR004794"  
/db\_xref="InterPro:IPR011549"  
/db\_xref="Uniprot/TREMBL:Q8B660"  
/translation="MMKKEDYMALALKEAEKMGFVAPNPLVGAVIVKDRRIISKGVSL  
KRFGDLHAERQAINKADEDISGSTLYVTLEPCCHVGKQPCTETALIKSGIKVVGSL  
DPNPLVSGKGIALLRKECLNVEGILREEDALNERFIFHTYKQPFVYLKYAMTLDG  
VCDSQLIPLDLSYLKVSAKTIPTWIATCSDNLAQQOQLKEMGRLIKVPRKGKLDLK  
VLMTILQEGIDSLILEGSSLHPSALKAGIVNRLIVFIAPKIIGGLAKAKTAISGEG  
DLNQDPRVKDIELSRMDSDWIEGKVEHYVYRN"  
1381..1387  
/gene="gbs0768"  
1381..1387  
/gene="gbs0768"  
1393..2043  
/gene="gbs0768"  
1393..2043  
/gene="gbs0768"  
/note="Similar to riboflavin synthase alpha chain"  
/codon\_start=1  
/transl\_table=11  
/product="Unknown"  
/protein\_id="CAD46412.1"  
/db\_xref="GI:23095202"  
/db\_xref="InterPro:IPR001793"  
/translation="MFTGIIEGMQVSRIRNGIKSQQLSIDAPKLVPLLRKGDSVAVN  
GVCLTVLDKSETAFIADVMPESMWRSLAALRLHLSKVLELALRSDRLRGHFLVGHV  
DGVKIEKIOKDDIAVRFSIDAPPSIMSYIIIEKGSVALDGLSLTVVSFTSHSPSVI  
PHTMAQNLSLKKVGDLLNIEVDVLGKYAEKFLAPTNRNTHTSVMDWSFLSENGY"  
2046..3254  
/gene="gbs0769"  
2046..3254  
/gene="gbs0769"  
2061..3254  
/gene="gbs0769"  
/note="Similar to GTP cyclohydrolase II/  
3,4-dihydroxy-2-butanone-4-phosphate synthase"  
/codon\_start=1  
/transl\_table=11  
/product="Unknown"  
/protein\_id="CAD46413.1"  
/db\_xref="GI:23095203"  
/db\_xref="GOA:Q8B658"  
/db\_xref="InterPro:IPR000422"  
/db\_xref="InterPro:IPR000926"  
/db\_xref="Uniprot/TREMBL:Q8B658"  
/translation="MAFSPYIKLLQDIKSGKVVVLMDDENRENEGDILCAEMVTKES  
INFMAKPGKGLICLPUSNYAEKLELAQMAHNTDNHETAPTISIDHLSTSTGISASD  
RALTKAKVNDSSKAKDFRPGHLPFLAKEGGVFLARNGHTATVLCRLAGLKECG  
CCETIMADGSMRKEDELLAPQKHDLAIAIKQDYRQREGGVVRREIQLPTQFG  
HFTAYGYSEVVANKEHVALVKGDISSGDEVLCRLHSECLTGDFVHSLRCDCEQLANA  
LQOIEAGRGVLLYMRQEGIGLINKLKAYHLQEEGLDLEANLALGFEGERDYG  
SAQLLDKLGINSINLLTNPNPKIQLEAEGICVKNRVPLQVAATAYDLNLYLTKYKEM  
GHLLD"  
3255..3739  
/gene="gbs0770"  
3255..3260  
/gene="gbs0770"  
3269..3739  
/gene="gbs0770"  
/note="Similar to riboflavin synthase complex beta chain"  
/codon\_start=1  
/transl\_table=11

terminator	gene	/product="Unknown"		/db_xref="UniProt/TREMBL:Q8E654"	
	CDS	/protein_id="CAD46414.1"		/translation="MKFYLVRHGKTDWNLNLEGRFGQANGSDSPLEAEIELEBEGQLVLSIHFDAVYSSDLGRADTVNILDANSCKPEIHVTPOLREWALGTLEGCKIATMQAIY	
		/db_xref="GI:23095204"		PROMTAFYQNPLOPKHDMFGAESLYQTHRVESPLRSLSKNDVKVLVGHGANLTAS	
		/db_xref="GOA:Q8E657"		IRSLGQYQSLHYKDKLDNASLTITETHDFKDFNCLTWNDSKSVLRQEVKMTW"	
		/db_xref="InterPro:IPR002180"		complement(7064..7069)	
		/db_xref="UniProt/Swiss-Prot:Q8E657"		/gene="gbs0773"	
		/db_xref="InterPro:IPR004364"		complement(7080..7553)	
		/db_xref="InterPro:IPR004365"		/gene="gbs0774"	
		/db_xref="InterPro:IPR006195"		complement(7080..7553)	
		/db_xref="InterPro:IPR008994"		/gene="gbs0774"	
terminator	gene	/note="Similar to putative phosphoglycerate mutase"		/note="Similar to transcriptional regulator"	
	CDS	/codon_start=1		/codon_start=1	
		/transl_table=11		/transl_table=11	
		/product="lysyl-tRNA synthetase"		Query Match	
		/protein_id="CAD46415.1"		Best Local Similarity 100.0%; Score 23; DB 1; Length 167050;	
		/db_xref="GI:23095205"		Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
		/db_xref="GOA:Q8E656"			
		/db_xref="InterPro:IPR002313"			
		/db_xref="InterPro:IPR004364"			
		/db_xref="InterPro:IPR004365"			
terminator	gene	/db_xref="InterPro:IPR006195"		Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23	
	CDS	/db_xref="InterPro:IPR006195"		Db 79838 TGAGAAGGCAGTAGAAGCTTAG 79860	
		/db_xref="InterPro:IPR008994"			
		/db_xref="UniProt/Swiss-Prot:Q8E656"			
		/translation="MSNOHIELNDQIIVREKMAALTEQGDIPGKGFERTATSGQL		RESULT 7	
		NEKYADSKEDLHDEETATTAGRLMTGRGKGVGFPAHQDREGQIQIYVRKDSVGE		CQ655069	
		NYEIPKADLGDGFVEQGVMTDMGELSIKATHILSKALRPLPEKPHGLTDIETI		LOCUS	
		YRKRHLDISNRDPRVTRKIIISRRFMDNSNGFLEVETPVLHNEAGSARPFPI		DEFINITION	
		THNQADIDMLRIATFELHKLRLVIGSERVIEIGRIIFNEGMDATHNPFETISIAVQ		CQ655069	
		AYADYQIMDLTEGIIQHVTKTVKGDGPINYPQTEIKINEPFRKRVHMDVAKEITGID		VERSION	
terminator	gene	FWKEMTEEAQAQAKENQVLEKHTFVGHIIINAFEEFVEDTIIQPTFVFGHVEVS		CQ655069.1 GI:41687946	
	CDS	PLAKNDYDPRFTDFELFTMTGKEYANAFTELNDPIQLSRFEAQAQAKELGDDEATG		SOURCE	
		VDYVTEALSYGMPETGGIGLIGDLRCMLLTDTTITRDVLLFPTMK"		ORGANISM	
		complement(5310..5315)		Streptococcus agalactiae	
		/gene="lysS"		Streptococcus agalactiae	
		5467..6411		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
		/gene="gbs0772"		Streptococcus	
		5467..5472		1	
		/gene="gbs0772"		REFERENCE	
		5479..6381		Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.	
terminator	gene	/note="Similar to unknown proteins"		and Tettein,H.	
	CDS	/codon_start=1		Nucleic acids and proteins from streptococcus groups a & b	
		/transl_table=11		Patent: WO 0234771-A 12026 02-MAY-2002;	
		/product="Unknown"		Chiron S.p.A. (It); THE INSTITUTE FOR GENOMIC RESEARCH (US)	
		/protein_id="CAD46416.1"		Location/Qualifiers	
		/db_xref="GI:23095206"		1..349980	
		/db_xref="GOA:Q8E655"		/organism="Streptococcus agalactiae"	
		/db_xref="InterPro:IPR005833"		/mol_type="unassigned DNA"	
		/db_xref="InterPro:IPR005834"		/db_xref="taxon:1311"	
		/db_xref="InterPro:IPR006439"		seq 10.967, from 0.000.001 to 0.349.980	
terminator	gene	/db_xref="UniProt/TREMBL:Q8E655"		seq 12.025, from 0.300.001 to 0.649.980	
	CDS	/translation="MTSIVFDVDDTIYDQQAQYRIAVEKCFPPDMSAINQAYIRFR		seq 12.026, from 0.600.001 to 0.949.980	
		HYSIDIGFPRVAGWETFEYFRWCKETLFEYREIDEATGIYQEIVHELENITM		seq 12.027, from 0.900.001 to 1.249.980	
		LDEMRTLDPLKSNVPMGIIITNGPTEHQKLVKVLGYDYVDPRVIVSQATGFKQP		seq 12.028, from 1.200.001 to 1.549.980	
		EKEIFNLAAEQDMPSTLYVGDSDINDMGAFNGWHSMFNRHRSLSKPKGIKPYI		seq 12.029, from 1.500.001 to 1.849.980	
		VDVTVLLRLTAQKEVLRMKYAR"		seq 12.030, from 1.800.001 to 2.149.980	
		6390..6411		seq 12.031, from 2.100.001 to 2.160.266"	
		/gene="gbs0772"			
		complement(6417..7069)			
		/gene="gbs0773"			
terminator	gene	complement(6417..7058)		Query Match	
	CDS	/gene="gbs0773"		Best Local Similarity 100.0%; Pred. No. 2.1;	
		/note="Similar to putative phosphoglycerate mutase"		Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
		/codon_start=1			
		/transl_table=11			
		/product="Unknown"		Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23	
		/protein_id="CAD46417.1"		Db 191615 TGAGAAGGCAGTAGAAGCTTAG 191637	
		/db_xref="GI:23095207"			
		/db_xref="GOA:Q8E654"			
		/db_xref="InterPro:IPR001345"			
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	
terminator	gene	AX954529		AX954529	
	CDS	LOCUS		Sequence 1375 from Patent WO03093306.	
		DEFINITION		ACCESSION	
		AX954529		VERSION	
		AX954529.1 GI:40783902		KEYWORDS	

SOURCE Streptococcus pyogenes  
ORGANISM Streptococcus pyogenes  
Bacteria; Firmicutes; Lactobacilliales; Streptococcaceae;  
Streptococcus.

REFERENCE 1  
AUTHORS Telford, J., Masignani, V., margarit y Ros, I., Grandi, G., Fraser, C. and Tettelin, H.  
TITLE Nucleic acids and proteins from streptococcus groups a b  
JOURNAL Patent: WO 03093306-A 1375 13-NOV-2003;  
Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
FEATURES Location/Qualifiers  
source 1. 349980  
/organism="Streptococcus pyogenes"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1314"  
/note="seq 1373 to long, original length: 2.160.266  
replaced by  
seq 1373: from 0.000.001 to 0.349.980  
seq 1374: from 0.300.001 to 0.649.980  
seq 1375: from 0.600.001 to 0.949.980  
seq 1376: from 0.900.001 to 1.249.980  
seq 1377: from 1.200.001 to 1.549.980  
seq 1378: from 1.500.001 to 1.849.980  
seq 1379: from 1.800.001 to 2.149.980  
seq 1380: from 2.100.001 to 2.160.266"

ORIGIN  
Query Match 100.0%; Score 23; DB 6; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23  
Db 220402 TGAGAGGCAGTAGAAGCTTAG 220424  
|||||

RESULT 9  
AC123147  
LOCUS AC123147.4 GI:24942304  
DEFINITION Rattus norvegicus clone CH230-333E19, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*  
AC123147  
VERSION AC123147.4  
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS\_ENRICHED.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 185291)  
Muzny, D. Marie., Metzker, M. Lee., Abranzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Centek, A., Chacko, J., Chavez, D., Chen, G., Chen, K., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Dexamco, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregiorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. B., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Li, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokemele, O., Okwunu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Pioppner, F., Poindexter, A., Popovic, D., Primus, E., Pu, L. L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojao, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Direct Submission  
Unpublished  
2 (bases 1 to 185291)  
Worley, K. C.  
Direct Submission  
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 185291)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 13, 2002 this sequence version replaced gi:23603979.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GMGN  
Center clone name: CH230-333E19  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 172302 bases at least Q40  
Consensus quality: 174607 bases at least Q30  
Consensus quality: 175894 bases at least Q20  
Estimated insert size: 181132; sum-of-contigs estimation  
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs

\* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 \* 1 185291: contig of 185291 bp in length.

## FEATURES

source  
 1..185291  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-333819"

## misc\_feature

1..1735  
 /note="wgs\_contig"

## misc\_feature

178867..179717  
 /note="clone\_boundary"

## misc\_feature

end sequence:B2234987"  
 183599..185291  
 /note="wgs\_end\_extension"  
 clone\_end:77"

## ORIGIN

Query Match 91.3%; Score 21; DB 14; Length 185291;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTGAAGCTTA 22

Db 59437 GAGAGGCGAGTGAAGCTTA 59457

## RESULT 10

AC121422/c

LOCUS

DEFINITION Rattus norvegicus clone CH230-11588, WORKING DRAFT SEQUENCE, 2

AC121422

AC121422.4 GI:25008025

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 231447)  
 Muzny, D., Marie, Metzker, M., Lee, A., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorensuewa, L., Loulseged, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Pal, L., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Rui, S., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaje, D., Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

## TITLE

Unpublished

2 (bases 1 to 231447)

Worley, K. C.

Direct Submission

Submitted (18-MAY-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 231447)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (15-NOV-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23196017.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Genome Center

Center: Baylor College of Medicine

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GXIS

Center clone name: CH230-11588

Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 211543 bases at least Q40

Consensus quality: 213418 bases at least Q30

Consensus quality: 214985 bases at least Q20

Estimated insert size: 215208; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).

\* NOTE: This is a 'working draft' sequence. It currently

consists of 2 contigs. The true order of the pieces

is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 227883: contig of 227883 bp in length

\* 227884 227983: gap of unknown length

\* 227984 231447: contig of 3464 bp in length.

#### FEATURES

source	Location/Qualifiers
	1..231447
	/organism="Rattus norvegicus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10116"
	/clone="CH230-115B8"
misc_feature	1..1431
	/note="wgs_end_extension"
	clone_end:Sp6"
misc_feature	2174..2753
	/note="clone_boundary"
	clone_end:Sp6"
	site:
misc_feature	end_sequence:BH292203"
	complement(221179..221702)
	/note="clone_boundary"
	clone_end:T7"
	site:
misc_feature	end_sequence:BH292202"
	224336..225446
	/note="wgs_end_extension"
	clone_end:T7"
misc_feature	226785..227883
	/note="wgs_end_extension"
	clone_end:T7"
gap	227884..227983
	/estimated_length=unknown
misc_feature	227984..229146
	/note="wgs_end_extension"
	clone_end:T7"
misc_feature	229860..231447
	/note="wgs_end_extension"
	clone_end:T7"
ORIGIN	
Query Match	91.3%; Score 21; DB 14; Length 231447;
Best Local Similarity	100.0%; Pred. No. 20;
Matches	21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	2 GAGAGGCGAGTAGAAGCTTA 22
Db	183282 GAGAGGCGAGTAGAAGCTTA 183262
RESULT 11	
AC095375	
LOCUS	AC095375 306616 bp DNA linear HTG 13-NOV-2002
DEFINITION	Rattus norvegicus clone CH230-2E1, WORKING DRAFT SEQUENCE, 4
	unordered pieces.
ACCESSION	AC095375
VERSION	AC095375.4 GI:24941434
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
	Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE	1 (bases 1 to 306616)
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alabrooks, S.L., Amaratunga, H.C., Are, J.R., Ayale, M., Banks, T., Barbaria, J., Benton, J., Blincoe, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buha, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,

Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earnhart, C., Edgar, D., Edwards, C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Haviak, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,  
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegged, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwunonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Fu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshtari, N., Sisson, I.,  
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.

Direct Submission  
 Unpublished  
 2 (bases 1 to 306616)  
 Worley, K.C.

Direct Submission  
 Submitted (16-SEP-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 306616)  
 Worley, K.C.

Direct Submission  
 Submitted (13-NOV-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Nov 13, 2002 this sequence version replaced gi:23264771.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
 in the feature table below represents a scaffold in the Atlas  
 assembly (a 'contig-scaffold'). Within each contig-scaffold,  
 individual sequence contigs are ordered and oriented, and separated  
 by sized gaps filled with Ns to the estimated size. The sequence  
 may extend beyond the ends of the clone and there may be sequence  
 contigs within a contig-scaffold that consist entirely of whole  
 genome shotgun sequence reads. Both end sequences and whole genome  
 shotgun sequence only contigs will be indicated in the feature  
 table.

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: http://www.hgsc.bcm.tmc.edu/  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 -----  
 Center project name: TUOF  
 Center clone name: CH230-2E1  
 ----- Summary Statistics  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 244944 bases at least Q40  
 Consensus quality: 247400 bases at least Q30  
 Consensus quality: 249174 bases at least Q20

Estimated insert size: 250113; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 260565: contig of 260565 bp in length  
\* 260566: gap of unknown length  
\* 260666: contig of 40417 bp in length  
\* 301083 301182: gap of unknown length  
\* 301183 302346: contig of 1164 bp in length  
\* 302347 302446: gap of unknown length  
\* 302447 306616: contig of 4170 bp in length.

## FEATURES

source  
1. .306616

/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-2E1"

misc\_feature

1. .1067  
/note="wgs\_end\_extension  
clone\_end:T7"

misc\_feature

4919..6420  
/note="wgs\_end\_extension  
clone\_end:T7"

misc\_feature

5053..5785  
/note="clone\_boundary  
clone\_end:T7  
site:"

misc\_feature

end\_sequence:BH262876"  
6620..7177  
/note="clone\_boundary  
clone\_end:Sp6  
site:"

gap

end\_sequence:BH262877"  
260566..260665  
/estimated\_length=unknown

misc\_feature

260666..261775  
/note="wgs\_end\_extension  
clone\_end:Sp6"

misc\_feature

26331..264680  
/note="wgs\_end\_extension  
clone\_end:Sp6"

gap

301083..301182  
/estimated\_length=unknown

gap

302347..302446  
/estimated\_length=unknown

## ORIGIN

Query Match 88.7%; Score 20.4; DB 14; Length 306616;  
Best Local Similarity 95.5%; Pred. No. 39;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGCTTAG 23

|||||  
DB 160683 GAGAGGCGAGTAGAAGCTTAG 160704

## RESULT 12

AR353664 AR353664 1378 bp DNA linear PAT 17-AUG-2003  
LOCUS  
DEFINITION Sequence 14 from patent US 6593093.  
ACCESSION AR353664  
VERSION AR353664.1 GI:33759697  
KEYWORDS

## SOURCE

ORGANISM

Unknown.

Unclassified.

1 (bases 1 to 1378)

Uhl,J.R. and Cockerill,F.R.

Detection of group a Streptococcus

Patent: US 6593093-A 14 15-JUL-2003;

JOURNAL Mayo Foundation for Medical Education and Research; Rochester, MN

FEATURES

source

1. .1378

/organism="unknown"

/mol\_type="genomic DNA"

## ORIGIN

Query Match 86.1%; Score 19.8; DB 6; Length 1378;

Best Local Similarity 91.3%; Pred. No. 1.2e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTTAG 23

|||||

9 TGAATATGCAGTAGAAGCTTAG 31

## RESULT 13

AX817901

LOCUS

AX817901

DEFINITION

Sequence 14 from Patent EP1338656.

ACCESSION

AX817901

VERSION

AX817901.1

KEYWORDS

GI:39723087

SOURCE

Streptococcus sp. 'group A'

Streptococcus sp. 'group A'

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Streptococcus.

REFERENCE

Uhl,J.R. and Cockerill Iii,F.R.

Detection of group a Streptococcus

Patent: EP 1338656-A 14 27-AUG-2003;

JOURNAL MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)

FEATURES

source

1. .1378

/organism="Streptococcus sp. 'group A'"

/mol\_type="unassigned DNA"

/db\_xref="taxon:36470"

/note="ptai sequence from isolate no. 10"

ORIGIN

Query Match 86.1%; Score 19.8; DB 6; Length 1378;

Best Local Similarity 91.3%; Pred. No. 1.2e+02;

Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTTAG 23

|||||

9 TGAATATGCAGTAGAAGCTTAG 31

## RESULT 14

AR353657

LOCUS

AR353657

DEFINITION

Sequence 7 from patent US 6593093.

ACCESSION

AR353657

VERSION

AR353657.1

KEYWORDS

GI:33759690

SOURCE

Unknown.

Unclassified.

1 (bases 1 to 1379)

Uhl,J.R. and Cockerill,F.R.

Detection of group a Streptococcus

Patent: US 6593093-A 7 15-JUL-2003;

JOURNAL Mayo Foundation for Medical Education and Research; Rochester, MN

FEATURES

source

1. .1379



ORIGIN	/organism="unknown"		/mol_type="genomic DNA"		
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2;
				Indels	0;
				Gaps	0;
QY	1	1	1	1	1
DB	9	9	9	9	9
RESULT 15	/organism="unknown"		/mol_type="genomic DNA"		
AXR353663	86.1%;		Score 19.8;	DB 6;	Length 1379;
LOCUS	Sequence 13 from patent US 6593093.		1379 bp	DNA	linear
DEFINITION	AR353663				
ACCESSION	AR353663				
VERSION	AR353663.1				
KEYWORDS	GI:33759696				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1379)				
AUTHORS	Uhl, J.R. and Cockerill, F.R.				
TITLE	Detection of group A Streptococcus				
JOURNAL	Patent: US 6593093-A 13 15-JUL-2003;				
FEATURES	Mayo Foundation for Medical Education and Research; Rochester, MN				
source	Location/Qualifiers				
	1..1379				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN	Unclassified.				
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2;
				Indels	0;
				Gaps	0;
QY	1	1	1	1	1
DB	9	9	9	9	9
RESULT 16	/organism="unknown"		/mol_type="genomic DNA"		
AXR353666	86.1%;		Score 19.8;	DB 6;	Length 1379;
LOCUS	Sequence 16 from patent US 6593093.		1379 bp	DNA	linear
DEFINITION	AR353666				
ACCESSION	AR353666				
VERSION	AR353666.1				
KEYWORDS	GI:33759699				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	1 (bases 1 to 1379)				
AUTHORS	Uhl, J.R. and Cockerill, F.R.				
TITLE	Detection of group A Streptococcus				
JOURNAL	Patent: US 6593093-A 16 15-JUL-2003;				
FEATURES	Mayo Foundation for Medical Education and Research; Rochester, MN				
source	Location/Qualifiers				
	1..1379				
	/organism="unknown"				
	/mol_type="genomic DNA"				
ORIGIN	Unclassified.				
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2;
				Indels	0;
				Gaps	0;
QY	1	1	1	1	1
DB	9	9	9	9	9
RESULT 17	/organism="unknown"		/mol_type="genomic DNA"		
AXR817894	86.1%;		Score 19.8;	DB 6;	Length 1379;
LOCUS	Sequence 7 from Patent EP1338656.		1379 bp	DNA	linear
DEFINITION	AXR817894				
ACCESSION	AXR817894				
VERSION	AXR817894.1				
KEYWORDS	GI:39723080				
SOURCE	Streptococcus sp. 'group A'				
ORGANISM	Streptococcus sp. 'group A'				
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
	Streptococcus.				
REFERENCE	1				
AUTHORS	Uhl, J.R. and Cockerill, F.R.				
TITLE	Detection of group A Streptococcus				
JOURNAL	Patent: EP 1338656-A 7 27-AUG-2003;				
FEATURES	MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)				
source	Location/Qualifiers				
	1..1379				
	/organism="Streptococcus sp. 'group A'"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:36470"				
	/note="ptai sequence from isolate no. 6"				
ORIGIN	Unclassified.				
Query Match	86.1%;	Score 19.8;	DB 6;	Length 1379;	
Best Local Similarity	91.3%;	Pred. No. 1.2e+02;			
Matches	21;	Conservative	0;	Mismatches	2;
				Indels	0;
				Gaps	0;
QY	1	1	1	1	1
DB	9	9	9	9	9
RESULT 18	/organism="unknown"		/mol_type="genomic DNA"		
AXR817900	86.1%;		Score 19.8;	DB 6;	Length 1379;
LOCUS	Sequence 13 from Patent EP1338656.		1379 bp	DNA	linear
DEFINITION	AXR817900				
ACCESSION	AXR817900				
VERSION	AXR817900.1				
KEYWORDS	GI:39723086				
SOURCE	Streptococcus sp. 'group A'				
ORGANISM	Streptococcus sp. 'group A'				
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
	Streptococcus.				
REFERENCE	1				
AUTHORS	Uhl, J.R. and Cockerill, F.R.				
TITLE	Detection of group A Streptococcus				
JOURNAL	Patent: EP 1338656-A 13 27-AUG-2003;				
FEATURES	MAYO FOUNDATION FOR MEDICAL EDUCATION				



[illegible]

```
Qy 1 TGAGAAGGCAGTAGAAAGCTTAG 23
Db 9 TGAATAATGCAGTAGAAAGCTTAG 31

RESULT 24
AR353659
LOCUS AR353659 1385 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 9 from patent US 6593093.
ACCESSION AR353659
VERSION AR353659.1 GI:33759692
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 1385)
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: US 6593093-A 9 15-JUL-2003;
MAYO Foundation for Medical Education and Research; Rochester, MN
FEATURES
  source
    1..1385
    /organism="unknown"
    /mol_type="genomic DNA"
ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1385;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAAGCTTAG 32

RESULT 25
AR353659
LOCUS AR353659 1385 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 9 from Patent EP1338656.
ACCESSION AR353659
VERSION AR353656.1 GI:39723082
KEYWORDS
SOURCE Streptococcus sp. 'group A'
ORGANISM Streptococcus sp. 'group A'
REFERENCE
  1
  Uhl,J.R. and Cockerill Iii,F.R.
  TITLE Detection of group A Streptococcus
  JOURNAL Patent: EP 1338656-A 9 27-AUG-2003;
  MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
  source
    1..1385
    /organism="Streptococcus sp. 'group A'"
    /mol_type="unassigned DNA"
    /db_xref="taxon:36470"
    /note="ptsi sequence from isolate no. 7"
ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1385;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAAGCTTAG 32

RESULT 26
AR353661
LOCUS AR353661 1390 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 11 from patent US 6593093.
ACCESSION AR353661
VERSION AR353665.1 GI:33759698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 1393)
  Uhl,J.R. and Cockerill,F.R.
  TITLE Detection of group A Streptococcus
  JOURNAL Patent: US 6593093-A 15 15-JUL-2003;
  MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; Rochester, MN
FEATURES
  source
    1..1390
    /organism="unknown"
    /mol_type="genomic DNA"
ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1390;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAAGCTTAG 23
Db 24 TGAATAATGCAGTAGAAAGCTTAG 46

RESULT 27
AR353661
LOCUS AR353661 1390 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 11 from Patent EP1338656.
ACCESSION AR353661
VERSION AR353665.1 GI:39723084
KEYWORDS
SOURCE Streptococcus sp. 'group A'
ORGANISM Streptococcus sp. 'group A'
REFERENCE
  1
  Uhl,J.R. and Cockerill Iii,F.R.
  TITLE Detection of group A Streptococcus
  JOURNAL Patent: EP 1338656-A 11 27-AUG-2003;
  MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
  source
    1..1390
    /organism="Streptococcus sp. 'group A'"
    /mol_type="unassigned DNA"
    /db_xref="taxon:36470"
    /note="ptsi sequence from isolate no. 9"
ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1390;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAAGCTTAG 23
Db 24 TGAATAATGCAGTAGAAAGCTTAG 46

RESULT 28
AR353665
LOCUS AR353665 1393 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 15 from patent US 6593093.
ACCESSION AR353665
VERSION AR353665.1 GI:33759698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
  1 (bases 1 to 1393)
  Uhl,J.R. and Cockerill,F.R.
  TITLE Detection of group A Streptococcus
  JOURNAL Patent: US 6593093-A 15 15-JUL-2003;
  MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; Rochester, MN
FEATURES
  source
    1..1390
    /organism="unknown"
    /mol_type="genomic DNA"
ORIGIN
Query Match 86.1%; Score 19.8; DB 6; Length 1390;
Best Local Similarity 91.3%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAAGCTTAG 23
Db 24 TGAATAATGCAGTAGAAAGCTTAG 46
```

```
ORIGIN
Query Match      86.1%;   Score 19.8; DB 6; Length 1447;
Best Local Similarity 91.3%; Pred. NO. 1.2e+02;
Matches 21: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```



```

/transl_table=11
/product="putative phosphoenolpyruvate:sugar
phosphotransferase system enzyme I"
/protein_id="AAK34199.1"
/db_xref="GI:13622482"
/translation="MTEMLKGTAAASDGVAVAKAYLLVQPDLSFETVTVDATNAEEARL
DVALQAQDELIVIRENAVESIGEEAARVFAHLMVLADPEMISQVKETIRAKQTNAE
TGLKVTDMFTITIPGMDNPMQBRADIRDAKRVLAHLGVKLPNPATINBESIV
TAHDLTPSDTALNQKQVKAFTVNIIGRTSHSAIMARTLEAAVLGTNDITKRVKDG
VIANGTIGEVILIDSEQVLAPEKAGAAAYAKQKAENSLDKAHTETADGKHFLAAN
IGTPKGVNDNGAEAVGLYKTEFLYMSDQDPTFEDQYEAIRKAVLGMGKPVVVR
TMIDGDELFPDIPKMNPLFLRALRISSETGDAMPTQRRALLRASVHGQLRI
NPFVALLAKFAAKAVDEEKANLLAEGVAVADDIQVGIMIETIPAAHAALADQFAKEV
DFPSTGTNDLIQYTWAAQDMNEQVSLYQVPNPSILRLNNVIAKAAHKGWAGMCGE
MAGDOQAVPLLVGMGLDFESNATSVLRTSLRSLKKLDSAKMEYANRALTECSAAEEV
LELSKEYVSD"
complement(3428..3691)
/gene="ptsh"
/locus_tag="SPy1373"
/locus_tag="SPy1373"
complement(3428..3691)
/gene="ptsh"
/locus_tag="SPy1373"
/notes="The N-terminal amino acid sequence of this ORF has
been determined from a spot isolated by 2-D gel
electrophoresis from another strain of S. pyogenes. Hogan,
D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen.
2000. Two-dimensional gel electrophoresis map of
Streptococcus pyogenes proteins. Unpublished data.; Best
Blastp hit = sp|P24366|P/HP STRSL PHOSPHOCARRIER PROTEIN
HPR (HISTIDINE-CONTAINING PROTEIN) >gi|480991|pir||S37585
phosphotransferase system phosphohistidine-containing
protein - Streptococcus salivarius
>gi|406779|emb|CAA78923.1| (Z17217) HPR [Streptococcus
salivarius] >gi|739971|prf||2004268A His containing
phosphocARRIER protein [Streptococcus salivarius]
>gi|1583333|prf||2120381A ptsh gene [Streptococcus
salivarius]"
/codon_start=1
/transl_table=11
/product="putative phosphotransferase system
phosphohistidine-containing protein"
/protein_id="AAK34200.1"
/db_xref="GI:13622483"
/translation="WASKDFHTVAETGTHARPATLLVOTASKFASDITLDYKGAVALN
KSIIMGVMSLGVQGGADVTISAGDAEDAIATIEETWKEGLA"
4084..4302
/gene="nrDH"
/locus_tag="SPy1374"
4084..4302
/gene="nrDH"
/locus_tag="SPy1374"
/notes="Best Blastp hit = sp|Q48708|NRDH LACLC
GLUTAREDOXIN-LIKE PROTEIN NRDH >gi|1222680|emb|CAA63372.1|
(X92690) glutaredoxin-like protein [Lactococcus lactis]"
/codon_start=1
/transl_table=11
/product="putative glutaredoxin"
/protein_id="AAK34201.1"
/db_xref="GI:13622484"
/translation="MTIVYSKNMCMQCKMTKKFLEHGQVNFQFINIDEHPEKVDYVKS
LGFTSAPVIEADNLVPSGQPAKLKELI"
4322..6481
/gene="nrDE.2"
/locus_tag="SPy1375"
4322..6481
/gene="nrDE.2"
/locus_tag="SPy1375"
/SC_number="1.17.4.1"
/notes="Best Blastp hit = gb|AAD41036.1|AF112535_3
(AF112535) ribonucleotide reductase alpha-chain
[Corynebacterium glutamicum]"
/codon_start=1
/transl_table=11
/product="putative ribonucleotide reductase alpha-chain"
/protein_id="AAK34202.1"
/db_xref="GI:13622485"
/translation="MSLKDLGDISFYRLNNEINRPVNGKIPLHKDKKEALKAFAENVL
PNTWSPTSIKEIYELISNDYIESAFIKYKPEPITELDSIIKSENPRFKFMAAYKF
YQVALKNDGEHYLENDRLVFNALTFADQGDLDKDLAVENINQRYQATPSFLN
AGRRERGVSCFLIQVTDNMSIGRSINSAQLSRVGGVGITLSNREAGAPIKGY
AGAAAGVVPVKNLFFSFSYGNQLQGGAGVYLVNVPDIIAFLSTKKNADEKVR
VKTLSLGIYVPDKELARKNEDMYLFSYVNEKEYGIPFNYLDTNNYDLVANPKI
TKTKIKARDLETEISKQOESGYPIINIDTANKANPIDGKIIMSNLCEILQVOTPS
LINDAQFVENGTDISNLGSTNLNMTSPDFGRSIKMTTRALTFTVDTSSIERAAPT
IKGNQAHPTLGAMGLHYLAQHHLIYSGSPESIEFTDIIYFMLLYNTLVSNNIAR
ERQTFVGFNGSKANGSYFDKYVYTHGFVPKSDLVKDLFKDHFIPQASDWSALRDVQ
KDYTHQVRLAFAPNQISYINDCSASIHPTQRIERQEKIKIYIYPANGSLDITI
PYTYSYMDMNRKVIDVYAAATEHVDQGLSLTLFLRSELMPELVEMTKQSQKQTTDRDS
ILRNVAENKGIKSIYIIRTFTDDGGEVGAQCSCVI"
6814..7773
/gene="nrdf.2"
/locus_tag="SPy1378"
6814..7773
/gene="nrdf.2"
/locus_tag="SPy1378"
/notes="Best Blastp hit = emb|CAA70766.1| (Y09572)
ribonucleotide reductase subunit R2F [Corynebacterium
ammoniaenes]"
/codon_start=1
/transl_table=11
/product="ribonucleotide diphosphate reductase small
subunit"
/protein_id="AAK34203.1"
/db_xref="GI:13622486"
/translation="MTTYEAINNNEIBDVIDKTSWEKLTQFWLDTIRIPSLNLDLDDW
KLSLQEKDLGVKFGGLTLLDTMOSETGVEAIRADVRTPHEEVALNNIQPMESVHAK
RYSLSQFLNTKKEIEEFWNTNNEFLQEKARIINDIVANGDALQKKVASTYLEFL
FYSFPFPLYVLGNKLANVAELIKLIIRDSVHGTYIGYKFGDLGNELPEDEQENFR
WMDYLLQLYENEKEETKTLTYDGVGWTEEVWTFURYANAKALMNLGQDPLPDTAND
VNPVIMGISTGTNHNHDFSGVNGYLGVSVEAMSDDDYNYGL"
7748..9061
/locus_tag="SPy1379"
7748..9061
/locus_tag="SPy1379"
/notes="Best Blastp hit = gb|AAF41386.1| (AE002449)
chloride channel protein-related protein [Neisseria
meningitidis MC58]"
/codon_start=1
Query Match 86.1%; Score 19.8; DB 1; Length 10430;
Best Local Similarity 91.3%; Pred. No. 1e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TCAGAGAGGCAGTAGAAGCTTAG 23
Db 3247 TGAATAATGAGTAGAAGCTTAG 3225
RESULT 36
AE010058/1 11992 bp DNA linear BCT 03-APR-2002
LOCUS Streptococcus pyogenes strain MGAS8232, section 106 of 173 of the
complete genome.
ACCESSION AE010058
VERSION AE010058.1 GI:19748527
KEYWORDS Streptococcus
SOURCE Streptococcus pyogenes MGAS8232
ORGANISM Streptococcus pyogenes MGAS8232
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 11992)
AUTHORS Smoot, J.C., Barbian, K.D., Van Gompel, J.J., Smoot, L.M.,
Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M.,
Porcella, S.F., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M.,
Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.G. and Musser, J.M.
Genome sequence and comparative microarray analysis of serotype M18

```

group A Streptococcus strains associated with acute rheumatic fever outbreaks  
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)  
11917108  
2 (bases 1 to 11992)  
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M., Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M., Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M., Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.George. and Musser,J.M.  
Direct Submission  
Submitted (31-JAN-2002) Laboratory of Human Bacterial Pathogenesis/Rocky Mountain Laboratories/NIH, 903 S. 4th St., Hamilton, MT 59840, USA

RTDFQIRKEFAQMKSQDSVQEQEVALDIPGEFDFLLDKINTIDD"  
complement (3016. .3513)  
/gene="spyM18\_1377"  
complement (3016. .3513)  
/genes="spyM18\_1377"  
/notes="best blastp match gb|AAK34192.1| (AE006574)  
conserved hypothetical protein [Streptococcus pyogenes M1 GAS]"

/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAL97972.1"  
/db\_xref="GI:19748531"  
/translation="MNKSKIEOVQLMIAQAKELFANESNALNLSASALLNMTLEN SVFTGYPEDGOELILGPFGQRCVCHIKLGKVCESASQSRRTILINDVKQANYIS CDAAMSEIVVPMVKEGHLIGVLDLSSLVADYDEVDOEYLETFFVLLLEKTYTFNM FGVKN"  
3658. .4467  
/gene="spyM18\_1378"  
3658. .4467  
/gene="spyM18\_1378"  
/notes="best blastp match gb|AAK34193.1| (AE006574)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL97973.1"  
/db\_xref="GI:19748532"  
/translation="MKQITQOKWLRYGLFIALVNLGIELLELGLTANNLSFKFAFALI LTLSLIGIYILPFAAAIFYLSKFFHNNLVIIIVSCLSGLYISGLSCGHLVQFPWS YIIPSKDALMGWDAPITAVEEPKASSAILVITLPPRLTKELVALLSGMGFOL TEIDRVLQAKSIDSLVPTAIGRISTAVTSHWVHTAFTIGAYLLKSGSLFSKQOQI FWLLSPLVHFIWNSPLTSIPGTVLLGTLILLIFGLDFQKINTLDDVLF"  
complement (4520. .4783)  
/gene="spyM18\_1379"  
complement (4520. .4783)  
/genes="spyM18\_1379"  
/notes="best blastp match gb|AAK34194.1| (AE006574)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL97974.1"  
/db\_xref="GI:19748533"  
/translation="MLLIWGLLFPVNNELFVVIYAPDKRAIKKRRISERNLLVITV LFGFGALLAAKXHHKTRKRWYFVITWYTSILTLVLYIFMK"  
complement (4863. .5489)  
/gene="udk"  
/note="synonym: spyM18\_1380"  
complement (4863. .5489)  
/gene="udk"  
/notes="best blastp match gb|AAK34195.1| (AE006574)  
putative uridine kinase [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative uridine kinase"  
/protein\_id="AAL97975.1"  
/db\_xref="GI:19748534"  
/translation="MFKKPIIIGVTGSGGKTSVSRALDSPNARIAMIQHDSYRK DQSHMSPFEERVKNYDHPADFDMIQQLKELLAGRPVDPIYDYKHKTRSNITFRQ DPQDVIIVEGILVLEDERLMDIKLFVDTDIDRIIRIRKDMMERGRSLESIIQD YTSVVRPMYHQFIEPSKRYADIVPEGVSNVVAIDVINSKIASILGEV"  
5587. .6672  
/gene="spyM18\_1381"  
5587. .6672  
/genes="spyM18\_1381"  
/notes="best blastp match gb|AAK34196.1| (AE006574)  
putative RNA helicase [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative RNA helicase"  
/protein\_id="AAL97976.1"  
/db\_xref="GI:19748535"

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

gene  
CDS

FEATURES  
source

gene

CDS

gene

CDS

gene

CDS

1. .11992  
/organism="Streptococcus pyogenes MGAS8232"  
/mol\_type="genomic DNA"  
/strain="MGAS8232"  
/db\_xref="taxon:186103"  
132\_..1073  
/gene="bira"  
/note="synonym: spyM18\_1374"  
132\_..1073  
/gene="bira"  
/note="best blastp match gb|AAK34189.1| (AE006574)  
putative biotin operon repressor [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative biotin operon repressor"  
/protein\_id="AAL97969.1"  
/db\_xref="GI:19748528"  
/translation="MKVSEKYYQLSQTTDFVSGEYLADQLSISRTSVWKSIXLENO GIQDLSUKHGRVQGDILIPKTIISGLGMPVITYTHSOSTQDADQGEAHNSAPR LYLAPQEAQGRLDQFASGTGSIYMSYLNKPNVPYAMPPTMVAASSIVKAISR LTGIDTETIWNVDIYGNHKGVLGTAITAVETGLITDVIIGVLNFFVTDPEAIA QKAGSLFTEKPTITRNLLIIDIWKLFLSIPVKDHVKVYKESVLNKNQVTFIENSQEK RAIADLTDQHLVQPDNGDLQTLRSGEISLSWEN"  
complement (1048. .1320)  
/gene="spyM18\_1375"  
complement (1048. .1320)  
/gene="spyM18\_1375"  
/notes="best blastp match gb|AAK34190.1| (AE006574)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL97970.1"  
/db\_xref="GI:19748529"  
/translation="MPLPTFYLVTFFHSTVFECYNKAWTIROFLFMAFVCAFEITYFN DLLLSGNLFAFLFWGLLFRDLRRVHTINQITKILKAANSFKKD"  
complement (1346. .3016)  
/gene="dnaX"  
/note="synonym: spyM18\_1376"  
complement (1346. .3016)  
/gene="dnaX"  
/notes="best blastp match gb|AAF98348.1| (AF280764) DNA polymerase III gamma/tau subunits [Streptococcus pyogenes]"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma/tau subunits"  
/protein\_id="AAL97971.1"  
/db\_xref="GI:19748530"  
/translation="MYQALYRKYSQTFDEMVGQSVISTTLKQAVESGKI SHAYLFSG PRGTGTSAAKIFAKAMNCPNQVGEPCNQCDICRIDTNGSLSDVIEDIASNGVDE IRDRKSTYAPSRATYKVYIIDVHMLSTGAFNALKTLEPTENVVFILATLELHK IPATILSRVORPEFKAIKQKAIREHLAWLDKEGIAVEVDALNLIARRAEGMEDALS ILDOALSLSPNQVITIAESITGISIGILALGGRVYVSQEQATQALAALETIYDSGK SMSRFADLLYLRDLVVKAGGNQKQSAVFDTNLSLSDIRIFQMIVTVTSHPLEIK KGTHPRIYAEMMTIQAQKEQILSQANLSGLISETITLNELQALQKQISQLQSRPD SLARSDTKPTGYRVDRTVILKIMEETVMSNQSQRYLDALKNWNEIIDNISAQD RALLMGSEPVLANSENAILAPEAAFNAEQVMNSRNNDMPFGNIMSKAAGFSFNILAVP

JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--







Db 107626 TGAAAAATGCAGTAGAAGCTTAG 107648

RESULT 39  
BA000034\_08  
WPCOMMENT

Sequence split into 19 fragments LOCUS BA000034 Accession BA000034

Fragment Name	Begin	End
BA000034_00	1	110000
BA000034_01	100001	210000
BA000034_02	200001	310000
BA000034_03	300001	410000
BA000034_04	400001	510000
BA000034_05	500001	610000
BA000034_06	600001	710000
BA000034_07	700001	810000
BA000034_08	800001	910000
BA000034_09	900001	1010000
BA000034_10	1000001	1110000
BA000034_11	1100001	1210000
BA000034_12	1200001	1310000
BA000034_13	1300001	1410000
BA000034_14	1400001	1510000
BA000034_15	1500001	1610000
BA000034_16	1600001	1710000
BA000034_17	1700001	1810000
BA000034_18	1800001	1894275

Continuation (9 of 19) of BA000034 from base 800001 (BA000034 Streptococcus pyogenes SS1

Query Match 86.1%; Score 19.8; DB 1; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 7626 TGAAAAATGCAGTAGAAGCTTAG 7648

RESULT 40  
CP000003\_10/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000003 Accession CP000003

Fragment Name	Begin	End
CP000003_00	1	110000
CP000003_01	100001	210000
CP000003_02	200001	310000
CP000003_03	300001	410000
CP000003_04	400001	510000
CP000003_05	500001	610000
CP000003_06	600001	710000
CP000003_07	700001	810000
CP000003_08	800001	910000
CP000003_09	900001	1010000
CP000003_10	1000001	1110000
CP000003_11	1100001	1210000
CP000003_12	1200001	1310000
CP000003_13	1300001	1410000
CP000003_14	1400001	1510000
CP000003_15	1500001	1610000
CP000003_16	1600001	1710000
CP000003_17	1700001	1810000
CP000003_18	1800001	1899877

Continuation (11 of 19) of CP000003 from base 1000001 (CP000003 Streptococcus pyogenes M

Query Match 86.1%; Score 19.8; DB 1; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 93181 TGAAAAATGCAGTAGAAGCTTAG 93159

RESULT 41  
CP0000017\_10/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP0000017 Accession CP0000017

Fragment Name	Begin	End
CP0000017_00	1	110000
CP0000017_01	100001	210000
CP0000017_02	200001	310000
CP0000017_03	300001	410000
CP0000017_04	400001	510000
CP0000017_05	500001	610000
CP0000017_06	600001	710000
CP0000017_07	700001	810000
CP0000017_08	800001	910000
CP0000017_09	900001	1010000
CP0000017_10	1000001	1110000
CP0000017_11	1100001	1210000
CP0000017_12	1200001	1310000
CP0000017_13	1300001	1410000
CP0000017_14	1400001	1510000
CP0000017_15	1500001	1610000
CP0000017_16	1600001	1710000
CP0000017_17	1700001	1810000
CP0000017_18	1800001	1838554

Continuation (11 of 19) of CP0000017 from base 1000001 (CP0000017 Streptococcus pyogenes 1

Query Match 86.1%; Score 19.8; DB 1; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 97184 TGAAAAATGCAGTAGAAGCTTAG 97162

RESULT 42  
CP0000056\_11/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP0000056 Accession CP0000056

Fragment Name	Begin	End
CP0000056_00	1	110000
CP0000056_01	100001	210000
CP0000056_02	200001	310000
CP0000056_03	300001	410000
CP0000056_04	400001	510000
CP0000056_05	500001	610000
CP0000056_06	600001	710000
CP0000056_07	700001	810000
CP0000056_08	800001	910000
CP0000056_09	900001	1010000
CP0000056_10	1000001	1110000
CP0000056_11	1100001	1210000
CP0000056_12	1200001	1310000
CP0000056_13	1300001	1410000
CP0000056_14	1400001	1510000
CP0000056_15	1500001	1610000
CP0000056_16	1600001	1710000
CP0000056_17	1700001	1810000
CP0000056_18	1800001	1897573

Continuation (12 of 19) of CP0000056 from base 1100001 (CP0000056 Streptococcus pyogenes 1

Query Match 86.1%; Score 19.8; DB 1; Length 110000;  
Best Local Similarity 91.3%; Pred. No. 84;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||

Db 20489 TGAAAAATGCAGTAGAAGCTTAG 20467

RESULT 43  
AC133332  
LOCUS

AC133332 140072 bp DNA linear HTG 10-SEP-2002

```

DEFINITION Felis catus clone RP86-331P23, WORKING DRAFT SEQUENCE, 4 unordered
pieces
AC133332
VERSION AC133332.1 GI:22773276
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Felis catus (cat)
ORGANISM Felis catus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Felidae;
Felinae; Felis.
1 (bases 1 to 140072)
Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C.,
Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J.,
Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
Laric,P., Lee-Lin,S.-Q., Legaapi,R., Maduro,Q.D., Maduro,V.B.,
Margulies,E.H., Mastello,C., Maskeri,B., Mastrian,S.D.,
McCluskey,J.C., McDowell,J., Paguirigan,C., Pearson,R.,
Pantnoy,M.E., Prasad,A., Reddix-Dugue,N., Schueler,M.G., Sison,C.,
Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 140072)
Green,E.D.
Direct Submission
Submitted (10-SEP-2002) NIH Intramural Sequencing Center, 8717
Groveomont Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoehngri.nih.gov
----- Project Information
Center project name: dao
Center clone name: 331P23
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 138836 bases at least Q40
Consensus quality: 138960 bases at least Q30
Consensus quality: 139106 bases at least Q20
Insert size: 111000; agarose-fp
Insert size: 139772; sum-of-contigs
Quality coverage: 12.82x in Q20 bases; agarose-fp
Quality coverage: 10.18x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2120: contig of 2120 bp in length
* 2121 2220: gap of unknown length
* 2221 33322: contig of 31102 bp in length
* 33323 33422: gap of unknown length
* 33423 77664: contig of 44242 bp in length
* 77665 77765: gap of unknown length
* 77765 140072: contig of 62308 bp in length.
Location/Qualifiers
1. .140072
/organism="Felis catus"
/mol_type="genomic DNA"
/db_xref="taxon:9685"
/clone="RP86-331P23"
/clone_lib="RP86"
1. .2120
/note="assembly_fragment"
2121. .2220

FEATURES
source
1. .140072
/note="estimated length=unknown"
2221. .33322
/note="assembly_fragment
clone_end:SP6
vector_side:right"
33323. .33422
/estimated length=unknown
33423. .77664
/note="assembly_fragment
clone_end:T7
vector_side:right"
77665. .77764
/estimated length=unknown
77765. .140072
/note="assembly_fragment"

misc_feature
86.1%; Score 19.8; DB 14; Length 140072;
Best Local Similarity 91.3%; Pred. No. 82;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TGAGAGCGCAGTAGAAAGCTTAG 23
| | | | | | | | | | | | | | | | | | | | | |
Db 17702 TAAGAGGCTAGTAGAAAGCTTAG 17724

RESULT 44
AC154824/c
LOCUS AC154824 Mus musculus BAC clone RP24-540A17 from chromosome 13, complete
sequence.
DEFINITION AC154824 Mus musculus BAC clone RP24-540A17 from chromosome 13, complete
sequence.
ACCESSION AC154824 AC133590
VERSION AC154824.2 GI:66773639
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 164494)
Adams,S. and Haglund,K.
The sequence of Mus musculus BAC clone RP24-540A17
Unpublished (2001)
2 (bases 1 to 164494)
Wilson,R.K.
Direct Submission
Submitted (30-DEC-2004) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 164494)
Wilson,R.K.
Direct Submission
Submitted (14-JAN-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
4 (bases 1 to 164494)
Wilson,R.K.
Direct Submission
Submitted (27-MAY-2005) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
5 (bases 1 to 164494)
Wilson,R.K.
Direct Submission
Submitted (21-JUN-2005) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On May 27, 2005 this sequence version replaced gi:56900450.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
----- Summary Statistics
Center project name: M_BB0540A17
Drafting center: WIBR

```



```
misc_feature      fragment_chain:1"
127158..179315
/notes="assembly_fragment:02008
fragment_chain:1"
179416..190326
/notes="assembly_fragment:00055
fragment_chain:1"
190427..193918
/notes="assembly_fragment:00018"
194019..203029
/notes="assembly_fragment:02581"
203130..227389
/notes="assembly_fragment:02583"
227490..270432
/notes="assembly_fragment:01065
clone_end:SP6
vector_side:right"

ORIGIN
Query Match      86.1%; Score 19.8; DB 14; Length 270432;
Best Local Similarity 91.3%; Pred.No. 78;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
||||| ||| ||||| |||||
Db 216240 TGAGAGACAGCAGAGAAAGCTTAG 216262

RESULT 46
CT010447/c
LOCUS
DEFINITION      CT010447 270432 bp DNA linear HTG 11-AUG-2005
MUS musculus chromosome 13 clone RP23-15K11, WORKING DRAFT
SEQUENCE, 10 unordered pieces.
CT010447
CT010447.2 GI:72534182
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
MUS musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 270432)
McLaren.S.
Direct Submission
Submitted (10-AUG-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Aug 12, 2005 this sequence version replaced gi:71142025.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
----- Project Information
Center project name: BM15K11
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 267600 bases at least Q40
Consensus quality: 268187 bases at least Q30
Consensus quality: 268483 bases at least Q20
Insert size: 269532; sum-of-contigs
Insert size: 250136; 4.8% error; agarose-fp
Quality coverage: 7.49x in Q20 bases; sum-of-contigs Quality
coverage: 8.20x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
```

```
* be preserved.
* 1 42623: contig of 42623 bp in length
* 42624 42723: gap of 100 bp
* 42724 76065: contig of 33342 bp in length
* 76066 76165: gap of 100 bp
* 76166 107675: contig of 31510 bp in length
* 107676 107776: gap of 101 bp
* 107777 127057: contig of 19281 bp in length
* 127058 127157: gap of 100 bp
* 127158 179315: contig of 52158 bp in length
* 179316 179415: gap of 100 bp
* 179416 190326: contig of 10911 bp in length
* 190327 190426: gap of 100 bp
* 190427 193918: contig of 3492 bp in length
* 193919 194019: gap of 101 bp
* 194020 203029: contig of 9010 bp in length
* 203030 203129: gap of 100 bp
* 203130 227389: contig of 24260 bp in length
* 227390 227489: gap of 100 bp
* 227490 270432: contig of 42943 bp in length.

FEATURES
            Location/Qualifiers
            1..270432
             /organism="Mus musculus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10090"
             /chromosome="13"
             /clone_lib="RPC1-23"
             /clone="RP23-15K11"
             1..42623
              /notes="assembly_fragment:01509
              fragment_chain:1"
              42724..76065
               /notes="assembly_fragment:00687
               fragment_chain:1"
               76166..107676
                /notes="assembly_fragment:00371
                fragment_chain:1"
                107777..127057
                 /notes="assembly_fragment:00182
                 fragment_chain:1"
                 127158..179315
                  /notes="assembly_fragment:02008
                  fragment_chain:1"
                  179416..190326
                   /notes="assembly_fragment:00055
                   fragment_chain:1"
                   190427..193918
                    /notes="assembly_fragment:02581"
                    194019..203029
                     /notes="assembly_fragment:02583"
                     203130..227389
                      /notes="assembly_fragment:01065
                      clone_end:SP6
                      vector_side:right"

ORIGIN
Query Match      86.1%; Score 19.8; DB 14; Length 270432;
Best Local Similarity 91.3%; Pred.No. 78;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
||||| ||| ||||| |||||
Db 238550 TGAGAGACAGCAGAGAAAGCTTAG 238528

RESULT 47
AC110439/c
LOCUS
DEFINITION      AC110439 319008 bp DNA linear HTG 11-OCT-2002
Rattus norvegicus clone CH230-287L21, *** SEQUENCING IN PROGRESS
***.
ACCESSION      AC110439
```

```

AC101439.4 GI:23604158
HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 319008)
REFERENCE
AUTHORS
Muzny,D,Marie., Metzker,M, Lee., Abranzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biewalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Greggorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,W.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jiang,H., Johnson,B., Johnson,R., Jollivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowig,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhsawa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Muidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plapper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puaoro,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,B., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 319008)
Worley,K.C.
Direct Submission
Submitted (13-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 319008)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 9, 2002 this sequence version replaced gi:21738108.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center of Medicine
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GQHE
Center Clone name: CH230-287L21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 169602 bases at least Q40
Consensus quality: 172273 bases at least Q30
Consensus quality: 173833 bases at least Q20
Estimated insert size: 175232; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces
is believed to be correct as given, however the sizes
of the gaps between them are based on estimates that have
been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 319008: contig of 319008 bp in length.
-----
FEATURES
source
1..319008
Location/Qualifiers
1..319008
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-287L21"
723..1636
/note="clone boundary
clone_end:Sp6
site:MboI
end sequence:RXAHS71TV"
complement(183223..184055)
/note="clone boundary
clone_end:T7
site:MboI
end sequence:RXAHS71TV"
184135..185293
/note="wgs_end_extension
clone_end:T7"
31971..319008
/note="wgs_end_extension
clone_end:T7"
-----
ORIGIN
Query Match 86.1%; Score 19.8; DB 14; Length 319008;
Best Local Similarity 91.3%; Pred. No. 77;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGAGAGGCAGCTAGAAAGCTTAG 23
DB 95734 TGAGAGGCAGCAGACAGCTTAG 95712
RESULT 48

```

```

AC009610      AC009610      60036 bp      DNA      linear      HTG 24-DEC-2000
LOCUS          Homo sapiens chromosome 4 clone RP11-3906 map 4, LOW-PASS SEQUENCE
DEFINITION     SAMPLING.
ACCESSION     AC009610
VERSION       AC009610.2 GI:11991318
KEYWORDS      HTG; HTGS_PHASE0.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
              Homnidae; Homo.
REFERENCE      1 (bases 1 to 60036)
AUTHORS       Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE         Homo sapiens chromosome 4, clone RP11-3906
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 60036)
AUTHORS       Birren,B., Linton,L., Barna,N., Beckerly,R., Benn,J., Brown,A.,
              Baker,J., Baldwin,J., Colangelo,M., Collins,S., Collymore,A.,
              Casle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
              Cooke,P., Dearellano,K., Depayre,S., Devon,K., Dewar,K.,
              Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
              Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
              Karas,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
              Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
              Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
              Meidrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
              Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
              Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
              Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
              Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
              Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE         Direct Submission
JOURNAL       Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT       On Dec 24, 2000 this sequence version replaced gi:5801798.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              ----- Project Information
              Project name: 12010
              Center clone name: 39_D_6
              -----
              * NOTE: This record contains 73 individual
              * sequencing reads that have not been assembled into
              * contigs. Runs of N are used to separate the reads
              * and the order in which they appear is completely
              * arbitrary. Low-pass sequence sampling is useful for
              * identifying clones that may be gene-rich and allows
              * overlap relationships among clones to be deduced.
              * However, it should not be assumed that this clone
              * will be sequenced to completion. In the event that
              * the record is updated, the accession number will
              * be preserved.
              *
              * 1 751: contig of 751 bp in length
              * 752 851: gap of 100 bp
              * 852 1590: contig of 739 bp in length
              * 1591 1690: gap of 100 bp
              * 1691 2381: contig of 691 bp in length
              * 2382 2481: gap of 100 bp
              * 2482 3211: contig of 730 bp in length
              * 3212 3311: gap of 100 bp
              * 3312 4041: contig of 730 bp in length
              * 4042 4141: gap of 100 bp
              * 4142 4890: contig of 749 bp in length
              * 4891 4991: gap of 100 bp
              * 4991 5702: contig of 712 bp in length
              * 5703 5802: gap of 100 bp
              *
              * 5803: contig of 701 bp in length
              * 6034: gap of 100 bp
              * 604 7296: contig of 693 bp in length
              * 7297 8124: contig of 728 bp in length
              * 8125 8224: gap of 100 bp
              * 8225 9035: contig of 711 bp in length
              * 9036 9765: contig of 730 bp in length
              * 9766 9865: gap of 100 bp
              * 9866 10606: contig of 741 bp in length
              * 10607 11426: contig of 720 bp in length
              * 11427 11526: gap of 100 bp
              * 11527 12256: contig of 730 bp in length
              * 12257 12356: gap of 100 bp
              * 12357 13108: contig of 752 bp in length
              * 13109 13208: gap of 100 bp
              * 13209 13921: contig of 713 bp in length
              * 13922 14021: gap of 100 bp
              * 14022 14738: contig of 717 bp in length
              * 14739 14838: gap of 100 bp
              * 14839 15537: contig of 699 bp in length
              * 15538 15637: gap of 100 bp
              * 15639 16374: contig of 737 bp in length
              * 16375 16474: gap of 100 bp
              * 16475 17211: contig of 737 bp in length
              * 17212 17311: gap of 100 bp
              * 17313 18043: contig of 732 bp in length
              * 18044 18143: gap of 100 bp
              * 18144 18853: contig of 710 bp in length
              * 18854 18953: gap of 100 bp
              * 18954 19686: contig of 733 bp in length
              * 19687 19786: gap of 100 bp
              * 19787 20515: contig of 729 bp in length
              * 20516 20615: gap of 100 bp
              * 20616 21367: contig of 752 bp in length
              * 21368 21467: gap of 100 bp
              * 21468 22219: contig of 752 bp in length
              * 22220 22319: gap of 100 bp
              * 22320 23100: contig of 681 bp in length
              * 23101 23100: gap of 100 bp
              * 23101 23826: contig of 726 bp in length
              * 23827 23926: gap of 100 bp
              * 23927 24618: contig of 692 bp in length
              * 24619 24718: gap of 100 bp
              * 24719 25453: contig of 735 bp in length
              * 25454 25553: gap of 100 bp
              * 25554 26267: contig of 714 bp in length
              * 26268 26367: gap of 100 bp
              * 26368 27108: contig of 741 bp in length
              * 27109 27208: gap of 100 bp
              * 27209 27930: contig of 722 bp in length
              * 27931 28030: gap of 100 bp
              * 28031 28743: contig of 713 bp in length
              * 28744 28843: gap of 100 bp
              * 28844 29574: contig of 731 bp in length
              * 29575 29674: gap of 100 bp
              * 29675 30399: contig of 725 bp in length
              * 30400 30499: gap of 100 bp
              * 30500 31265: contig of 766 bp in length
              * 31266 31365: gap of 100 bp
              * 31366 32077: contig of 712 bp in length
              * 32078 32177: gap of 100 bp
              * 32178 32901: contig of 724 bp in length
              * 32902 33001: gap of 100 bp
              * 33002 33733: contig of 732 bp in length
              * 33734 33833: gap of 100 bp
              * 33834 34529: contig of 696 bp in length
              * 34530 34629: gap of 100 bp
              * 34630 35378: contig of 749 bp in length
              * 35379 35478: gap of 100 bp
              * 35479 36171: contig of 693 bp in length

```







TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Homo sapiens chromosome 4, clone RP11-39D6  
Unpublished  
2 (bases 1 to 164084)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G.,  
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domino, M., Doyle, M., Ginde, S., Goyette, M., Graham, L.,  
Galan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McSwan, P., McGurk, A., McKernan, K., McPheters, R.,  
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Oct 20, 2001 this sequence version replaced gi:7284115.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE  
JOURNAL  
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L8360  
Center clone name: 39\_D\_6  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 154712 bases at least Q40  
Consensus quality: 159436 bases at least Q30  
Consensus quality: 161272 bases at least Q20  
Insert size: 157000; agarose-fp  
Insert size: 162484; sum-of-contigs  
Quality coverage: 4.9 in Q20 bases; agarose-fp  
Quality coverage: 4.7 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1  
\* 6107: contig of 6107 bp in length  
\* 6108  
\* 6207: gap of 100 bp  
\* 6208  
\* 7644: contig of 1437 bp in length  
\* 7645  
\* 7744: gap of 100 bp  
\* 7745  
\* 11040: contig of 3296 bp in length  
\* 11041  
\* 11140: gap of 100 bp  
\* 11141  
\* 12843: contig of 1702 bp in length  
\* 12843  
\* 12943: gap of 100 bp  
\* 12943  
\* 16834: contig of 3892 bp in length  
\* 16835  
\* 16934: gap of 100 bp  
\* 16935  
\* 20681: contig of 3747 bp in length  
\* 20682  
\* 20781: gap of 100 bp  
\* 20782  
\* 25004: contig of 4223 bp in length  
\* 25005  
\* 25104: gap of 100 bp

FEATURES  
source

1. 164084  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-39D6"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 6107  
/note="assembly\_fragment"  
clone\_end:SP6  
vector\_side:left  
6108..6207  
/estimated\_length=100  
6208..7644  
/note="assembly\_fragment"  
7645..7744  
/estimated\_length=100  
7745..11040  
/note="assembly\_fragment"  
11041..11140  
/estimated\_length=100  
11141..12842  
/note="assembly\_fragment"  
12843..12942  
/estimated\_length=100  
12943..16834  
/note="assembly\_fragment"  
16835..16934  
/estimated\_length=100  
16935..20681  
/note="assembly\_fragment"  
20682..20781  
/estimated\_length=100  
20782..25004  
/note="assembly\_fragment"  
25005..25104  
/estimated\_length=100  
25105..31308  
/note="assembly\_fragment"  
31309..31408  
/estimated\_length=100  
31409..37851  
/note="assembly\_fragment"  
37852..37951  
/estimated\_length=100  
37952..44358  
/note="assembly\_fragment"  
44359..44458  
/estimated\_length=100  
44459..52433  
/note="assembly\_fragment"  
52434..52533

\* 25105 31308: contig of 6204 bp in length  
\* 31309 31408: gap of 100 bp  
\* 31409 37851: contig of 6443 bp in length  
\* 37852 37951: gap of 100 bp  
\* 37952 44358: contig of 6407 bp in length  
\* 44359 52433: contig of 7975 bp in length  
\* 52434 52533: gap of 100 bp  
\* 52534 58556: contig of 6023 bp in length  
\* 58557 58556: gap of 100 bp  
\* 58557 72273: contig of 13617 bp in length  
\* 72274 72373: gap of 100 bp  
\* 72374 87291: contig of 14918 bp in length  
\* 87292 87391: gap of 100 bp  
\* 87392 106498: contig of 19107 bp in length  
\* 106499 106598: gap of 100 bp  
\* 106599 144410: contig of 37812 bp in length  
\* 144411 144510: gap of 100 bp  
\* 144511 164084: contig of 19574 bp in length.



Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuehwa, L., Loulseghe, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwakwileme, O., Okwono, G., Olarnpungsoo, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plommer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Vallas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

Direct Submission  
Unpublished  
2 (bases 1 to 19119)  
Worley, K.C.  
Submitted (21-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 19119)  
Rat Genome Sequencing Consortium.  
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 19, 2002 this sequence version replaced gi:23907920.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

Center project name: GWYH  
Center clone name: CH230-30903  
Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 176412 bases at least Q40

Consensus quality: 179015 bases at least Q30  
Consensus quality: 180610 bases at least Q20  
Estimated insert size: 181867; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length (see <http://www.hgsc.bcm.tmc.edu/docs/Genbankdraftdata.html>).  
NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 189365: contig of 189365 bp in length  
189366 189465: gap of unknown length  
189466 191119: contig of 1654 bp in length.

FEATURES  
source

1. 191119  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clones="CH230-30903"  
150921..179194  
/note="clone boundary"  
clone\_end:Sp5  
site:  
end\_sequence:BZ124947"  
159049..160115  
/note="wgs end extension"  
clone\_end:Sp6"  
189365..189465  
/estimated\_length=unknown

misc\_feature

gap

ORIGIN

Query Match 81.7%; Score 18.8; DB 14; Length 191119;  
Best Local Similarity 90.9%; Pred. No. 2.5e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAAGCTTAG 23

Db 61950 GAGAGGCGAGTAGAAAGCATGG 61929

RESULT 53

AC117689

LOCUS

DEFINITION

AC117689

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 200562)

Birren, B., Linton, L., Nuebaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

AC117689 200562 bp DNA linear ROD 17-JAN-2004  
Mus musculus chromosome 12, clone RP23-446119, complete sequence.

AC117689 11 GI:41014873

HTG.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 200562)

Birren, B., Nuebaum, C. and Lander, E.

Mus musculus chromosome 12, clone RP23-446119

Unpublished

2 (bases 1 to 200562)

Birren, B., Linton, L., Nuebaum, C., Lander, E., Ali, A., Allen, N.,

Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,

Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,

Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A.,

Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J.S., Dodge, S.,

Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,

Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,

Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,

Kamat, A., Karatas, A., Kelle, C., LaRocque, K., Lamazares, R.,

Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,

Center clone name: 446\_I\_19

-----

FEATURES

source

Location/Qualifiers

1. 200562

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/chromosome="12"

/map="12"

/clone="RP23-446119"

/clone\_lib="RPCI-23 Female Mouse BAC"

misc\_feature

1. 569

/note="wgs\_end\_extension"

clone\_end:SP6"

570. 575

/note="clone boundary"

clone\_end:SP6

site:Boori"

615. 972

/rpt\_family="MER76"

complement(1851. 1927)

/rpt\_family="ID\_B1"

2002. 2040

/rpt\_family="AT\_rich"

2904. 3075

/rpt\_family="L1MA7"

complement(3406. 9520)

/rpt\_family="L1P"

9542. 9579

/rpt\_family="AT\_rich"

9858. 9896

/rpt\_family="TAAAAA"

9989. 10209

/rpt\_family="B4A"

complement(10668. 11339)

/rpt\_family="Lx2"

11365. 11527

/rpt\_family="Lx2"

complement(11531. 11709)

/rpt\_family="L1\_MM"

11710. 12632

/rpt\_family="Lx2"

12671. 12699

/rpt\_family="CAA"

complement(13503. 13652)

/rpt\_family="MERS8"

14354. 15308

/rpt\_family="RLTRL13C"

15335. 15366

/rpt\_family="TTTGTG"

15991. 16308

/rpt\_family="Lx9"

16963. 16985

/rpt\_family="CAAA"

17157. 18470

/rpt\_family="Lx9"

18856. 18987

/rpt\_family="GGAA"

19564. 19611

/rpt\_family="TA"

20821. 20864

/rpt\_family="CAAAAA"

complement(20936. 21231)

/rpt\_family="L1\_MM"

21233. 21388

/rpt\_family="L1\_MM"

21844. 21889

/rpt\_family="TTTGTG"

complement(22130. 22306)

/rpt\_family="ID\_B1"

complement(23225. 23466)

/rpt\_family="URRIA"

complement(24053. 24218)

McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (10-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 200562)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (09-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 200562)

Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., DeArellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., MacLean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (17-JAN-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jan 17, 2004 this sequence version replaced gi:39573811.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L24076

-----

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

```

/rpt family="Lx2"
2421g .24988
/rpt family="Lx"
25227 .31255
/rpt family="L1_MM"
31398 .31426
/rpt family="(TC)n"
3142g .31468
/rpt family="(TA)n"
31580 .31651
/rpt family="(T)n"
complement(31661 .31798)
/rpt family="B1_MM"
complement(31877 32017)
/rpt family="URR1B"
32031 .32052
/rpt family="AT_rich"
32201 .32300
/rpt family="(T)n"
33621 .33681
/rpt family="(CA)n"
complement(33745 .33880)
/rpt family="B4g"
33954 .33982
/rpt family="AT_rich"
3416g .35945
/rpt family="Lx2"
36660 .36695

Query Match      81.7%; Score 18.8; DB 9; Length 200562;
Best Local Similarity 90.9%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGGAGGCGAGTAGAAGCTTA 22
|||||
Db 140449 TGAGGAGGCGAGTAGAAGCTTA 140470

RESULT 54
AC107536/c
LOCUS
DEFINITION
AC107536 239438 bp DNA linear HTG 13-MAY-2003
Rattus norvegicus clone CH230-211L16, *** SEQUENCING IN PROGRESS
***, 3 unordered pieces.
AC107536
VERSION
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 239438)
Muzny,D,Marie., Metzker,M, Lee., Abramson, S., Amin, A., Anguiano, D.,
Allen, C., Allen, H., Alsbrooks, S., Ayodeji, M., Baca, E., Baden, H.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Barber, M., Barnstead, M., Benahmed, F.,
Baldwin, D., Bandaranaike, D., Barber, M., Blyth, P., Brown, M.,
Biswalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpthy, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensuhewa, L., Loulseged, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Margum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwaokemele, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Primus, E., Pu, L.-L.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, B., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S.J., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
Unpublished
2 (bases 1 to 239438)
Worley, K.C.
Direct Submission
Submitted (23-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 239438)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On May 13, 2003 this sequence version replaced gi:24942470.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/) Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKPU
Center clone name: CH230-211L16
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 21615 bases at least Q40
Consensus quality: 21613 bases at least Q30
Consensus quality: 21897 bases at least Q20
Estimated insert size: 226897; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 .5696: contig of 5696 bp in length
* 5697 .5796: gap of unknown length
* 5797 .238072: contig of 232176 bp in length
* 237973 .238072: gap of unknown length
* 238073 .239438: contig of 1366 bp in length.
FEATURES             Location/Qualifiers
     source            1..239438
                        /organism="Rattus norvegicus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10116"
                        /clones="CH230-211L16"
     misc_feature      1..1924
                        /note="wgs_end_extension
                        clone_end:Sp6"
                        1975..5696
                        /note="wgs_end_extension
                        clone_end:Sp6"
                        complement(2994..3834)
                        /note="clone_boundary
                        clone_end:Sp6
                        site:EcoRI
                        end_sequence:RWBLD68TVB"
                        5697..5796
                        /estimated_length=unknown
                        5797..7308
                        /note="wgs_contig"
     misc_feature      169595..172710
                        /note="wgs_contig"
     misc_feature      189620..191165
                        /note="wgs_contig"
                        193829..195622
                        /note="wgs_contig"
                        complement(230967..231795)
                        /note="clone_boundary
                        clone_end:T7
                        site:EcoRI
                        end_sequence:RWBLD68TVB"
                        237973..238072
                        /estimated_length=unknown
ORIGIN
Query Match          81.7%; Score 18.8; DB 14; Length 239438;
Best Local Similarity 90.9%; Pred. No. 2.4e-02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy      2 GAGAGGCGAGTAGAAGGCTTAG 23
      |||||
Db      90230 GAGAGGCGAGTAGAAGGCTTGTG 90209
      |||||

RESULT 55
AC128575/c
LOCUS      AC128575             239808 bp    DNA    linear    HTG 12-OCT-2002
DEFINITION Rattus norvegicus clone CH230-149J17, WORKING DRAFT SEQUENCE, 3
            unordered pieces.
ACCESSION  AC128575
VERSION     AC128575.2  GI:23907831
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
REFERENCE  1 (bases 1 to 239808)
AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,

```

```

Allen,C., Allen,H., Alsbrooks,S., Amin,A., Auguiano,D.,
Ayalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Behanmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroil,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Gants,R., Garcia,A., Garner,T., Garza,M.,
Georgiades,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheehwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Napua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Rizly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojasa,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajd,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villanueva,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 239808)
Worley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 239808)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (12-OCT-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 12, 2002 this sequence version replaced gi:21909348.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome

```

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
-----  
Center project name: GROQ  
Center clone name: CH230-149J17  
----- Summary Statistics

Assembly program: Phrap; version 0.990329  
Consensus quality: 166535 bases at least Q40  
Consensus quality: 168913 bases at least Q30  
Consensus quality: 170484 bases at least Q20  
Estimated insert size: 168276; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 237163: contig of 237163 bp in length  
\* 237164 237263: gap of unknown length  
\* 237264 238491: contig of 1228 bp in length  
\* 238492 238592: gap of unknown length  
\* 238592 239808: contig of 1217 bp in length.

## FEATURES

source

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

gap

gap

## ORIGIN

Query Match 81.7%; Score 18.8; DB 14; Length 239808;  
Best Local Similarity 90.9%; Pred. No. 2.4e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 GAGAGGCGACTAGAAAGCTTAG 23  
|||||  
DB 184147 GAGAGGCGACTAGAAAGCATGG 184126

## RESULT 56

AC128356

LOCUS

DEFINITION

AC128356

AC128356.3 GI:25007899

HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.

HTG; HTGS PHASE1; HTGS DRAFT; HTGS FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 243280)

Murphy D. Marie., Metzker M. Lee., Abramson S., Adams C., Alder J.,

Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,

Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,

Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,

Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,

Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,

Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,

Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen J.,

Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,

Davila M., Davis C., Davy-Carroll L., De Anda C., Dederich D.,

Delgado O., Denson S., Dezamo C., Ding Y., Dinh H., Divya K.,

Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,

Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,

Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,

Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,

Georgievski S., Geer K., Gill R., Grady M., Guerra W., Guevara W.,

Gunaratne P., Haaland W., Hamill C., Hamilton C., Hamilton K.,

Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,

Hernandez R., Hines S., Hladun S.L., Hodgson A., Hogues M.,

Hollins B., Howells S., Hulyk S., Hume J., Idlbird D., Jackson A.,

Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jollivet A.,

Karpathy S., Kelly S., Khan Z., King L., Kovar C.,

Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,

Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,

Lorensheva L., Louisedge H., Lozano R.J., Lu X., Ma J.,

Mareshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,

Mangum B., Mapua P., Martin K., Martin R., Martinez E.,

Mawhiney S., McLeod M.P., McNeill T.Z., Meenen E.,

Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,

Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Naif L.,

Nankervis C., Neal D., Newton N., Nguyen N., Norris S., Parks K.,

Nwakoelameh O., Okwuonu G., Olarnpunsagoon A., Pal S., Parks K.,

Pasternak S., Paul H., Perez A., Perez L., Pfannkuch C.,

Plopper F., Poindexter A., Popovic D., Primus E., Pu L.-L.,

Puazo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,

Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs P.,

Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,

Sanders W., Savary G., Scherer S., Scott G., Shatsman S., Shen H.,

Shetty J., Shvartsbeyn A., Sison I., Sitter C.D., Smajls D.,

Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,

Steinle M., Strong R., Sutton A., Svatek A., Tabor P., Taylor C.,

Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,

Valas R., Vera V., Villasana D., Waldron L., Walker B., Wang J.,

Wang Q., Wang S., Warren J., Warren R., Wei X., White F.,

Williams G., Willson R., Wlezyk R., Woodden H., Worley K.,

Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,

Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von

Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,

Weinstock G. and Gibbs R.A.

Direct Submission

Unpublished

2 (bases 1 to 243280)

Worley K.C.

Direct Submission

Submitted (19-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 243280)

REFERENCE



```

AUTHORS      Rat Genome Sequencing Consortium.
TITLE        Direct Submission
JOURNAL      Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
              of Molecular and Human Genetics, Baylor College of Medicine, One
              Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Nov 15, 2002 this sequence version replaced gi:23195600.
              The sequence in this assembly is a combination of BAC based reads
              and whole genome shotgun sequencing reads assembled using Atlas
              (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
              in the feature table below represents a scaffold in the Atlas
              assembly (a 'contig-scaffold'). Within each contig-scaffold,
              individual sequence contigs are ordered and oriented, and separated
              by sized gaps filled with Ns to the estimated and oriented. and sequence
              may extend beyond the ends of the clone and there may be sequence
              contigs within a contig-scaffold that consist entirely of whole
              genome shotgun sequence reads. Both end sequences and whole genome
              shotgun sequence only contigs will be indicated in the feature
              table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZCQ
Center clone name: CH230-146115
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 234015 bases at least Q40
Consensus quality: 236096 bases at least Q30
Consensus quality: 237191 bases at least Q20
Estimated insert size: 237240; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 5270: contig of 5270 bp in length
* 5271 5370: gap of unknown length
* 5371 240402: contig of 235032 bp in length
* 240403 240502: gap of unknown length
* 240503 241700: contig of 1198 bp in length
* 241701 241800: gap of unknown length
* 241801 243280: contig of 1480 bp in length.
*
* Location/Qualifiers
*     1..243280
*         /organism="Rattus norvegicus"
*         /mol_type="genomic DNA"
*         /db_xref="taxon:10116"
*         /clone="CH230-146115"
*     1..1867
*         /note="wgs end extension
* clone_end:T7"
* 1918..5270
*     /note="wgs_end_extension
* clone_end:T7"
* 5271..5370
*     /estimated_length=unknown
* 5371..6611
*     /note="wgs_end_extension
* clone_end:T7"
* 8802..10545
*     /note="wgs_end_extension
* clone_end:T7"
* 10200..10384
*     /note="clone_boundary

misc_feature
source      .
            /organism="Rattus norvegicus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10116"
            /clone="CH230-146115"
            1..1867
            /note="wgs end extension
clone_end:T7"

misc_feature
1918..5270
    /note="wgs_end_extension
clone_end:T7"

gap
5271..5370
    /estimated_length=unknown

misc_feature
5371..6611
    /note="wgs_end_extension
clone_end:T7"

misc_feature
8802..10545
    /note="wgs_end_extension
clone_end:T7"

misc_feature
10200..10384
    /note="clone_boundary

clone_end:T7
site:
end sequence:RWBAE56TJB"
236702..240402
    /note="wgs contig"
240403..240502
    /estimated_length=unknown
241701..241800
    /estimated_length=unknown

ORIGIN
Query Match      81.7%; Score 18.8; DB 14; Length 243280;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAGGCGAGTAGAAGCTTAG 23
    |||||
Db 85443 GAGAGGCGAGTAGAAGGCTTG 85464

RESULT 57
HSL81824/c
LOCUS      HSL81824
DEFINITION Homo sapiens (subclone 2_a12 from P1 H39) DNA sequence, complete
sequence.
ACCESSION  L81824
VERSION     L81824.1 GI:1930410
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 3310)
            Martin,C.H., Arcaina,T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomocan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
            Sequencing of human chromosome 5q
            Unpublished (1996)
            2 (bases 1 to 3310)
            Martin,C.H., Arcaina,T., Bondoc,M.M., Chiang,A., Critz,P.A.,
            Davis,C.A., Doyle,C.M., Ericsson,C.L., Farfan,D.E., Gunning,K.M.,
            Houston,K.A., Jaklevic,M.A., Kadner,K.E., Kim,K., Kim,S.F.,
            Ko,C.L., Lewis,K.D., Li,M., Lindquist,K.J., Lomocan,M.A.,
            Lustre,V.M., Machrus,M.U., Mayeda,C.A., Miguel,T.M., Miller,C.A.,
            Mok,M.S., Pacleb,J.M., Patel,S.G., Santos,R.F., Subramanian,S.,
            Wan,K.H., Whitelaw,K.R., Yee,A., Yeh,R.T., Yu,C. and Palazzolo,M.J.
            Direct Submission
            Submitted (09-APR-1997)
            Original source text: Homo sapiens (Subclones in pOT2 from P1 clone
            H39) DNA.
            Sequence submitted by:
            Human Genome Center
            Lawrence Berkeley National Laboratory, MS 74-157
            Berkeley, CA 94720
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Website (http://www.hgc.lbl.gov/sequence-archive.html) or
            send email to human@genome.lbl.gov.
            Location/Qualifiers
            1..3310
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="P1 H39 (5236)"
            /sub_clone="2_a12"

FEATURES
source
1..3310
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="5"
    /clone="P1 H39 (5236)"
    /sub_clone="2_a12"

ORIGIN

```



```

Query Match      80.0%; Score 18.4; DB 8; Length 3310;
Best Local Similarity 95.0%; Pred. No. 5.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCT 20
    |||||
Db 2348 TAAGAAGGCAGTAGAAGCT 2329
    |||||

RESULT 58
AC009010 Homo sapiens chromosome 5 clone XXp1-1352A1, WORKING DRAFT
LOCUS SEQUENCE, 8 unordered pieces.
DEFINITION AC009010
AC009010.5 GI:11178047
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 60634)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 60634)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Nov 15, 2000 this sequence version replaced gi:9964734.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1189015, H231
Center clone name: XXp1-1352A1
-----
Summary Statistics
Consensus quality: 50043 bases at least Q40
Consensus quality: 55954 bases at least Q30
Consensus quality: 57601 bases at least Q20
Estimated insert size: 8000; pulse field gel estimation
Quality coverage: 2.82 in Q20 bases; pulse field gel estimation
Quality coverage: 3.77 in Q20 bases; sun-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 1297: contig of 1297 bp in length
* 1298 1397: gap of unknown length
* 1398 2981: contig of 1584 bp in length
* 2982 3081: gap of unknown length
* 3082 5483: contig of 2402 bp in length
* 5484 5583: gap of unknown length
* 5584 11308: contig of 5725 bp in length
* 11309 11408: gap of unknown length
* 11409 16459: contig of 5051 bp in length
* 16460 16559: gap of unknown length
* 16560 25749: contig of 9190 bp in length
* 25750 25849: gap of unknown length
* 25850 34850: contig of 9010 bp in length
* 34860 34959: gap of unknown length
* 34960 60634: contig of 25675 bp in length.
Location/Qualifiers

FEATURES
source
1..60791
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="XXp1-1352A1"
1298..1397
/estimated_length=unknown
2982..3081
/estimated_length=unknown
5484..5583
/estimated_length=unknown
11309..11408
/estimated_length=unknown
16460..16559
/estimated_length=unknown
25750..25849
/estimated_length=unknown
34860..34959
/estimated_length=unknown

ORIGIN
Query Match      80.0%; Score 18.4; DB 14; Length 60634;
Best Local Similarity 95.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCT 20
    |||||
Db 35406 TAAGAAGGCAGTAGAAGCT 35425
    |||||

RESULT 59
AC004749/c Homo sapiens chromosome 5, P1 clone 632A8 (LBNL H29), complete
LOCUS sequence.
DEFINITION AC004749 AC000974 AC000975 AC000977 AC001489 AC002208
L48484 L49048 L77054 L77055 L77056 L77057 L78758 L78759
L81379 L81380 L81628 L81795 L81796 L81797 L81798
AC004749.1 GI:3157918
VERSION HTG.
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
REFERENCE 1 (bases 1 to 60791)
AUTHORS Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE Sequencing of human chromosome 5
JOURNAL Unpublished
JOURNAL 2 (bases 1 to 60791)
REFERENCE Ricke, D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
JOURNAL 3 (bases 1 to 60791)
REFERENCE Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,
Davis, C.A., Kadner, K., Miguel, T., Pitluck, S., Pollard, M.,
Rojeski, H., Subramanian, S. and Martin, C.H.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
Location/Qualifiers
1..60791
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"

```

```
repeat_region /map="5c"
repeat_region /clone="f32a8"
repeat_region /note="DBNL H29"
repeat_region 1..60791
repeat_region /note="(T) 27"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit="t"
repeat_region 453..757
repeat_region /rpt_family="MER1B"
misc_feature complement(1157..1283)
misc_feature /note="GRAIL 2 excellent exon, frame 0"
misc_feature complement(1177..1282)
misc_feature /note="83% identity mouse mp17d11.r1"
repeat_region /db_xref="dbEST:AA288684"
repeat_region complement(1882..2044)
repeat_region /rpt_family="Alu"
repeat_region 1891..1913
repeat_region /note="(T) 23"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit="t"
repeat_region 5204..5503
repeat_region /rpt_family="Alu"
repeat_region 6376..6912
repeat_region /rpt_family="Alu"
misc_feature 8028..8115
repeat_region /note="GRAIL 2 excellent exon, frame 2"
repeat_region complement(9142..9288)
repeat_region /rpt_family="MER5"
repeat_region 9356..9674
repeat_region /rpt_family="Alu"
repeat_region 9611..10326
repeat_region /rpt_family="L1"
repeat_region complement(10569..10869)
misc_feature /rpt_family="Alu"
misc_feature 10859..10943
misc_feature /note="GRAIL 2 excellent exon, frame 1"
repeat_region complement(11133..11417)
repeat_region /note="99% identity EST60905"
repeat_region /db_xref="dbEST:AA352848"
repeat_region complement(12263..12538)
repeat_region /rpt_family="Alu"
repeat_region 12276..12396
repeat_region /note="(T) 21"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit="t"
repeat_region complement(14099..14403)
repeat_region /rpt_family="Alu"
misc_feature 14892..15199
misc_feature /note="97% identity EST02082"
repeat_region /db_xref="dbEST:M85566"
repeat_region 16203..16362
repeat_region /rpt_family="Alu"
repeat_region 17932..18067
repeat_region /rpt_family="MER42"
repeat_region complement(18549..18857)
repeat_region /rpt_family="Alu"
repeat_region 18552..18868
repeat_region /note="(T) 21"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit="t"
repeat_region complement(19792..20110)
repeat_region /rpt_family="Alu"
repeat_region complement(20826..21152)
misc_feature 21162..21549
misc_feature /note="99% identity yu10e03.r1"
repeat_region /db_xref="dbEST:H79907"
repeat_region complement(21500..21824)
repeat_region /rpt_family="Alu"
repeat_region 22829..23118
repeat_region /rpt_family="Alu"
repeat_region 23149..23457
repeat_region /rpt_family="Alu"
```

---

```
repeat_region complement(23659..23973)
repeat_region /rpt_family="Alu"
repeat_region 24620..25456
repeat_region /rpt_family="LTR1"
repeat_region 25651..26710
repeat_region /rpt_family="Alu"
repeat_region 25933..26038
repeat_region /rpt_family="Alu"
repeat_region 26390..28695
repeat_region /rpt_family="Alu"
repeat_region complement(26796..26904)
repeat_region /rpt_family="MIR"
misc_feature complement(27505..27616)
misc_feature /note="99% identity"
repeat_region /db_xref="dbEST:D63193"
repeat_region complement(29907..30188)
repeat_region /rpt_family="Alu"
repeat_region complement(31287..31532)
repeat_region /rpt_family="Alu"
repeat_region 31649..31828
repeat_region /rpt_family="MER46"
repeat_region 31831..31976
repeat_region /rpt_family="Alu"
repeat_region complement(32558..32880)
repeat_region /rpt_family="Alu"
repeat_region complement(32887..33189)
repeat_region /rpt_family="Alu"
repeat_region 33219..33558
repeat_region /rpt_family="Alu"
repeat_region 33559..33843
repeat_region /rpt_family="Alu"
repeat_region complement(34684..34783)
repeat_region /rpt_family="L1"
repeat_region 35259..35558
repeat_region /rpt_family="Alu"
repeat_region 35561..35867
repeat_region /rpt_family="Alu"
repeat_region 36176..36203
repeat_region /note="(AC) 14"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit="ac"
repeat_region complement(37312..42768)
repeat_region /rpt_family="L1"
repeat_region complement(43077..43389)
repeat_region /rpt_family="Alu"
repeat_region 43834..44199
repeat_region /rpt_family="Alu"
repeat_region 44128..44446
repeat_region /rpt_family="Alu"
repeat_region 44424..44448
repeat_region /note="(A) 20"
repeat_region /rpt_type=tandem
repeat_region /rpt_unit="a"
repeat_region complement(44448..44837)
repeat_region /rpt_family="Alu"
repeat_region complement(44895..45025)
repeat_region /rpt_family="Alu"
repeat_region 45828..46140
repeat_region /rpt_family="Alu"
repeat_region complement(46222..46303)
repeat_region /rpt_family="MER42C"
repeat_region 46466..46636
repeat_region /rpt_family="MER44C"
repeat_region complement(46987..47304)
repeat_region /rpt_family="Alu"
repeat_region complement(47536..47862)
repeat_region /rpt_family="Alu"
repeat_region 48111..48231
repeat_region /rpt_family="MIR"
repeat_region 48705..54852
repeat_region /rpt_family="L1"
repeat_region complement(55742..56085)
repeat_region /rpt_family="Alu"
```

Search completed: January 27, 2006, 23:13:39  
Job time : 915.685 secs

```
repeat_region 55940..55959
/note="(T)20"
/rpt_type=tandem
/rpt_unit="t"
repeat_region complement(56109..56261)
/rpt_family="MER20"
misc_feature complement(56361..56800)
/note="97% identity yp96g11.r1"
/db_xref="dbEST:R92125"
repeat_region 56771..56875
/rpt_family="MLT1F"
repeat_region complement(56884..57190)
/rpt_family="Alu"
repeat_region 58783..58908

Query Match 80.0%; Score 18.4; DB 8; Length 60791;
Best Local Similarity 95.0%; Pred. No. 4.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCT 20
DB 20410 TAAGAGGCGAGTAGAAGCT 20391

RESULT 60
AC009016 79041 bp DNA linear PRI 08-NOV-2000
LOCUS Homo sapiens chromosome 5 clone P1_889E7, complete sequence.
DEFINITION AC009016
ACCESSION AC009016
VERSION AC009016.3 GI:11120759
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 79041)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 79041)
DOE Joint Genome Institute.
Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS 3 (bases 1 to 79041)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
JOURNAL Submitted (08-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 8, 2000 this sequence version replaced gi:6862871.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
FEATURES
source
1..79041
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="5"
/clone="P1_889E7"

ORIGIN
Query Match 80.0%; Score 18.4; DB 8; Length 79041;
Best Local Similarity 95.0%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCT 20
DB 65068 TAAGAGGCGAGTAGAAGCT 65087
```

**This Page Blank (uspto)**



```
QY 1 TGCATGTATGGGTTATCTTCC 21
Db 1 TGCATGTATGGGTTATCTTCC 21

RESULT 2
LOCUS AR353653 21 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 2 from patent US 6593093.
ACCESSION AR353653
VERSION AR353653.1 GI:33759686
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 21)
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: US 6593093-A 2 15-JUL-2003;
MAYO Foundation for Medical Education and Research; Rochester, MN
FEATURES
source
1..21
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 1 TGCATGTATGGGTTATCTTCC 21

RESULT 3
AX817889
LOCUS AX817889 21 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 2 from Patent EPI338656.
ACCESSION AX817889
VERSION AX817889.1 GI:39723076
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE 1
AUTHORS Uhl,J.R. and cockerill Iii,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 2 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
source
1..21
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xrefs="taxon:32830"
/note="Oligonucleotide"
ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 21;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 1 TGCATGTATGGGTTATCTTCC 21

RESULT 4
AX353662/c
LOCUS AR353662 1323 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 12 from patent US 6593093.
ACCESSION AR353662
```

```
VERSION AR353662.1 GI:33759695
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1323)
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: US 6593093-A 12 15-JUL-2003;
MAYO Foundation for Medical Education and Research; Rochester, MN
FEATURES
source
1..1323
Location/Qualifiers
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1323;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 193 TGCATGTATGGGTTATCTTCC 173

RESULT 5
AX817899/c
LOCUS AX817899 1323 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 12 from Patent EPI338656.
ACCESSION AX817899
VERSION AX817899.1 GI:39723085
KEYWORDS
SOURCE Streptococcus sp. 'group A'
ORGANISM Streptococcus sp. 'group A'
Bacteria; Firmicutes; Lactobacilliales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Uhl,J.R. and cockerill Iii,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 12 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
source
1..1323
Location/Qualifiers
/organism="Streptococcus sp. 'group A'"
/mol_type="unassigned DNA"
/db_xrefs="taxon:36470"
/note="ptai sequence from isolate no. 2"
ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1323;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 193 TGCATGTATGGGTTATCTTCC 173

RESULT 6
AX353664/c
LOCUS AR353664 1378 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 14 from patent US 6593093.
ACCESSION AR353664
VERSION AR353664.1 GI:33759697
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1378)
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: US 6593093-A 14 15-JUL-2003;
MAYO Foundation for Medical Education and Research; Rochester, MN
```

```
FEATURES
  source
    Location/Qualifiers
      1..1378
      /organism="unknown"
      /mol_type="genomic DNA"
ORIGIN
  Query Match
    100.0%; Score 21; DB 6; Length 1378;
  Best Local Similarity
    100.0%; Pred. No. 11;
  Matches
    21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy
  1 TGCATGTATGGGTTATCTTCC 21
  |||||
Db
  209 TGCATGTATGGGTTATCTTCC 189

RESULT 7
AX817901/c
LOCUS
  AX817901 1378 bp DNA linear PAT 10-DEC-2003
DEFINITION
  Sequence 14 from Patent EPI338656.
ACCESSION
  AX817901
VERSION
  AX817901.1 GI:39723087
KEYWORDS
  Streptococcus sp. 'group A'
SOURCE
  Streptococcus sp. 'group A'
  Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
  Streptococcus.
REFERENCE
  1 Uhl, J.R. and cockerill Iii, F.R.
  AUTHORS
  Uhl, J.R. and cockerill Iii, F.R.
  TITLE
  Detection of group A Streptococcus
  JOURNAL
  Patent: EP 1338656-A 14 27-AUG-2003;
  MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
  source
    Location/Qualifiers
      1..1378
      /organism="Streptococcus sp. 'group A'"
      /mol_type="unassigned DNA"
      /db_xref="taxon:36470"
      /note="ptai sequence from isolate no. 10"
ORIGIN
  Query Match
    100.0%; Score 21; DB 6; Length 1378;
  Best Local Similarity
    100.0%; Pred. No. 11;
  Matches
    21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy
  1 TGCATGTATGGGTTATCTTCC 21
  |||||
Db
  209 TGCATGTATGGGTTATCTTCC 189

RESULT 8
AR353657/c
LOCUS
  AR353657 1379 bp DNA linear PAT 17-AUG-2003
DEFINITION
  Sequence 7 from patent US 6593093.
ACCESSION
  AR353657
VERSION
  AR353657.1 GI:33759690
KEYWORDS
  Unknown.
SOURCE
  Unknown.
REFERENCE
  1 (bases 1 to 1379)
  Uhl, J.R. and Cockerill, F.R.
  TITLE
  Detection of group A Streptococcus
  JOURNAL
  Patent: US 6593093-A 7 15-JUL-2003;
  MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; ROCHESTER, MN
FEATURES
  source
    Location/Qualifiers
      1..1379
      /organism="unknown"
      /mol_type="genomic DNA"
ORIGIN
  Query Match
    100.0%; Score 21; DB 6; Length 1379;
  Best Local Similarity
    100.0%; Pred. No. 11;
  Matches
    21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy
  1 TGCATGTATGGGTTATCTTCC 21
  |||||
Db
  209 TGCATGTATGGGTTATCTTCC 189

RESULT 9
AR353663/c
LOCUS
  AR353663 1379 bp DNA linear PAT 17-AUG-2003
DEFINITION
  Sequence 13 from patent US 6593093.
ACCESSION
  AR353663
VERSION
  AR353663.1 GI:33759696
KEYWORDS
  Unknown.
SOURCE
  Unknown.
REFERENCE
  1 (bases 1 to 1379)
  Uhl, J.R. and Cockerill, F.R.
  TITLE
  Detection of group A Streptococcus
  JOURNAL
  Patent: US 6593093-A 13 15-JUL-2003;
  MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; ROCHESTER, MN
FEATURES
  source
    Location/Qualifiers
      1..1379
      /organism="unknown"
      /mol_type="genomic DNA"
ORIGIN
  Query Match
    100.0%; Score 21; DB 6; Length 1379;
  Best Local Similarity
    100.0%; Pred. No. 11;
  Matches
    21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy
  1 TGCATGTATGGGTTATCTTCC 21
  |||||
Db
  209 TGCATGTATGGGTTATCTTCC 189

RESULT 10
AR353666/c
LOCUS
  AR353666 1379 bp DNA linear PAT 17-AUG-2003
DEFINITION
  Sequence 16 from patent US 6593093.
ACCESSION
  AR353666
VERSION
  AR353666.1 GI:33759699
KEYWORDS
  Unknown.
SOURCE
  Unknown.
REFERENCE
  1 (bases 1 to 1379)
  Uhl, J.R. and Cockerill, F.R.
  TITLE
  Detection of group A Streptococcus
  JOURNAL
  Patent: US 6593093-A 16 15-JUL-2003;
  MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; ROCHESTER, MN
FEATURES
  source
    Location/Qualifiers
      1..1379
      /organism="unknown"
      /mol_type="genomic DNA"
ORIGIN
  Query Match
    100.0%; Score 21; DB 6; Length 1379;
  Best Local Similarity
    100.0%; Pred. No. 11;
  Matches
    21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy
  1 TGCATGTATGGGTTATCTTCC 21
  |||||
Db
  210 TGCATGTATGGGTTATCTTCC 190

RESULT 11
AX817894/c
LOCUS
  AX817894 1379 bp DNA linear PAT 10-DEC-2003
DEFINITION
  Sequence 7 from Patent EPI338656.
ACCESSION
  AX817894
VERSION
  AX817894.1 GI:39723080
KEYWORDS
  Streptococcus sp. 'group A'
SOURCE
```



```

ORGANISM      Streptococcus sp. 'group A'
               Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE      1
AUTHORS        Uhl,J.R. and cockerill Iii,F.R.
TITLE          Detection of group A Streptococcus
JOURNAL        Patent: EP 1338656-A 7 27-AUG-2003;
               MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES       Location/Qualifiers
               1..1379
               /organism="Streptococcus sp. 'group A'"
               /mol_type="unassigned DNA"
               /db_xref="taxon:36470"
               /note="ptSI sequence from isolate no. 6"
ORIGIN
Query Match   100.0%; Score 21; DB 6; Length 1379;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||
Db      209 TGCATGTATGGGTTATCTTCC 189

RESULT 12
AX817900/c
LOCUS      AX817900               1379 bp      DNA      linear      PAT 10-DEC-2003
DEFINITION Sequence 13 from Patent EP1338656.
ACCESSION  AX817900
VERSION     AX817900.1 GI:39723086
KEYWORDS   Streptococcus sp. 'group A'
SOURCE      Streptococcus sp. 'group A'
ORGANISM    Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE   1
AUTHORS     Uhl,J.R. and cockerill Iii,F.R.
TITLE       Detection of group A Streptococcus
JOURNAL     Patent: EP 1338656-A 13 27-AUG-2003;
             MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES    Location/Qualifiers
             1..1379
             /organism="Streptococcus sp. 'group A'"
             /mol_type="unassigned DNA"
             /db_xref="taxon:36470"
             /note="ptSI sequence from isolate no. 4"
ORIGIN
Query Match   100.0%; Score 21; DB 6; Length 1379;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||
Db      209 TGCATGTATGGGTTATCTTCC 189

RESULT 13
AX817903/c
LOCUS      AX817903               1379 bp      DNA      linear      PAT 10-DEC-2003
DEFINITION Sequence 16 from Patent EP1338656.
ACCESSION  AX817903
VERSION     AX817903.1 GI:39723089
KEYWORDS   Streptococcus sp. 'group A'
SOURCE      Streptococcus sp. 'group A'
ORGANISM    Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
             Streptococcus.
REFERENCE   1
AUTHORS     Uhl,J.R. and cockerill Iii,F.R.
TITLE       Detection of group A Streptococcus
JOURNAL     Patent: EP 1338656-A 16 27-AUG-2003;
             MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES    Location/Qualifiers
             1..1379
             /organism="Streptococcus sp. 'group A'"
             /mol_type="unassigned DNA"
             /db_xref="taxon:36470"
             /note="ptSI sequence from isolate no. 4"
ORIGIN
Query Match   100.0%; Score 21; DB 6; Length 1379;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||
Db      209 TGCATGTATGGGTTATCTTCC 189

RESULT 14
AR353658/c
LOCUS      AR353658               1384 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 8 from patent US 6593093.
ACCESSION  AR353658
VERSION     AR353658.1 GI:33759691
KEYWORDS   Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 1384)
AUTHORS     Uhl,J.R. and Cockerill,F.R.
TITLE       Detection of group A Streptococcus
JOURNAL     Patent: US 6593093-A 8 15-JUL-2003;
             MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; Rochester, MN
FEATURES    Location/Qualifiers
             1..1384
             /organism="unknown"
             /mol_type="genomic DNA"
ORIGIN
Query Match   100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||
Db      210 TGCATGTATGGGTTATCTTCC 190

RESULT 15
AR353660/c
LOCUS      AR353660               1384 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 10 from patent US 6593093.
ACCESSION  AR353660
VERSION     AR353660.1 GI:33759693
KEYWORDS   Unknown.
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 1384)
AUTHORS     Uhl,J.R. and Cockerill,F.R.
TITLE       Detection of group A Streptococcus
JOURNAL     Patent: US 6593093-A 10 15-JUL-2003;
             MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH; Rochester, MN
FEATURES    Location/Qualifiers
             1..1384
             /organism="unknown"
             /mol_type="genomic DNA"
ORIGIN
Query Match   100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
        |||||||
Db      210 TGCATGTATGGGTTATCTTCC 190

```

```
Qy 1 TGCATGATGGGTTATCTTCC 21
Db 209 TGCATGATGGGTTATCTTCC 189

RESULT 16
AX817895/c
LOCUS AX817895 1384 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 8 from Patent EP1338656.
ACCESSION AX817895
VERSION AX817895.1 GI:39723081
KEYWORDS Streptococcus sp. 'group A'
SOURCE Streptococcus sp. 'group A'
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1
AUTHORS Uhl, J.R. and cockerill Iii, F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 8 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
    Location/Qualifiers
    1..1384
    /organism="Streptococcus sp. 'group A'"
    /mol_type="unassigned DNA"
    /db_xref="taxon:36470"
    /note="ptse sequence from isolate no. 5"
source

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 210 TGCATGATGGGTTATCTTCC 190

RESULT 17
AX817897/c
LOCUS AX817897 1384 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 10 from Patent EP1338656.
ACCESSION AX817897
VERSION AX817897.1 GI:39723083
KEYWORDS Streptococcus sp. 'group A'
SOURCE Streptococcus sp. 'group A'
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1
AUTHORS Uhl, J.R. and cockerill Iii, F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 10 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
    Location/Qualifiers
    1..1384
    /organism="Streptococcus sp. 'group A'"
    /mol_type="unassigned DNA"
    /db_xref="taxon:36470"
    /note="ptse sequence from isolate no. 8"
source

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 209 TGCATGATGGGTTATCTTCC 189

RESULT 18
AX817895/c
LOCUS AX817895 1384 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 8 from Patent EP1338656.
ACCESSION AX817895
VERSION AX817895.1 GI:39723081
KEYWORDS Streptococcus sp. 'group A'
SOURCE Streptococcus sp. 'group A'
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1
AUTHORS Uhl, J.R. and cockerill Iii, F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 8 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
    Location/Qualifiers
    1..1384
    /organism="Streptococcus sp. 'group A'"
    /mol_type="unassigned DNA"
    /db_xref="taxon:36470"
    /note="ptse sequence from isolate no. 5"
source

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1384;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 210 TGCATGATGGGTTATCTTCC 190

RESULT 19
AX817896/c
LOCUS AX817896 1385 bp DNA linear PAT 10-DEC-2003
DEFINITION Sequence 9 from Patent EP1338656.
ACCESSION AX817896
VERSION AX817896.1 GI:39723082
KEYWORDS Streptococcus sp. 'group A'
SOURCE Streptococcus sp. 'group A'
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1
AUTHORS Uhl, J.R. and cockerill Iii, F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 9 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
    Location/Qualifiers
    1..1385
    /organism="Streptococcus sp. 'group A'"
    /mol_type="unassigned DNA"
    /db_xref="taxon:36470"
    /note="ptse sequence from isolate no. 7"
source

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1385;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTTATCTTCC 21
Db 210 TGCATGATGGGTTATCTTCC 190

RESULT 20
AX817896/c
LOCUS AX817896 1390 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 11 from patent US 6593093.
ACCESSION AX817896
VERSION AX817896.1 GI:33759694
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1390)
AUTHORS Uhl, J.R. and Cockerill, F.R.
```

TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 11 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN

FEATURES  
source  
1. .1390  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 1390;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 224 TGCATGTATGGGTTATCTTCC 204

RESULT 21  
LOCUS AX817898/c  
DEFINITION Sequence 11 from Patent EPI338656.  
ACCESSION AX817898  
VERSION AX817898.1 GI:39723084  
KEYWORDS Streptococcus sp. 'group A',  
SOURCE Streptococcus sp. 'group A',  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE 1  
AUTHORS Uhl,J.R. and cockerill Iii,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: EP 1338656-A 11 27-AUG-2003;  
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)  
FEATURES  
source  
1. .1390  
/organism="Streptococcus sp. 'group A'"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:36470"  
/note="ptSI sequence from isolate no. 9"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 1390;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 224 TGCATGTATGGGTTATCTTCC 204

RESULT 22  
LOCUS AR353665/c  
DEFINITION Sequence 15 from patent US 6593093.  
ACCESSION AR353665  
VERSION AR353665.1 GI:33759698  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1393)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 15 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
1. .1393  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 1393;

Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 23  
LOCUS AX817902/c  
DEFINITION Sequence 15 from Patent EPI338656.  
ACCESSION AX817902  
VERSION AX817902.1 GI:39723088  
KEYWORDS Streptococcus sp. 'group A',  
SOURCE Streptococcus sp. 'group A',  
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE 1  
AUTHORS Uhl,J.R. and cockerill Iii,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: EP 1338656-A 15 27-AUG-2003;  
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)  
FEATURES  
source  
1. .1393  
/organism="Streptococcus sp. 'group A'"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:36470"  
/note="ptSI sequence from isolate no. 3"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 1393;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 24  
LOCUS AR353667/c  
DEFINITION Sequence 17 from patent US 6593093.  
ACCESSION AR353667  
VERSION AR353667.1 GI:33759700  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1447)  
AUTHORS Uhl,J.R. and Cockerill,F.R.  
TITLE Detection of group a Streptococcus  
JOURNAL Patent: US 6593093-A 17 15-JUL-2003;  
Mayo Foundation for Medical Education and Research; Rochester, MN  
FEATURES  
source  
1. .1447  
/organism="unknown"  
/mol\_type="genomic DNA"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 1447;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 25  
LOCUS AX817904/c

```
LOCUS AX817904 1447 bp DNA linear PAT 10-DRC-2003
DEFINITION Sequence 17 from Patent EP1338656.
ACCESSION AX817904
VERSION AX817904.1 GI:39723090
SOURCE Streptococcus sp. 'group A'
ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Uhl,J.R. and cockerill iii,F.R.
TITLE Detection of group A Streptococcus
JOURNAL Patent: EP 1338656-A 17 27-AUG-2003;
MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)
FEATURES
source
Location/Qualifiers
1..1447
/organism="Streptococcus sp. 'group A'"
/mol_type="unassigned DNA"
/db_xref="taxon:36470"
/notes="pfsi sequence from isolate no. 11"
ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1447;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTATCTTCC 21
|||||
Db 210 TGCATGTATGGGTATCTTCC 190
|||||
RESULT 26
CO647508/c
LOCUS CO647508 1731 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 4465 from Patent WO234771.
ACCESSION CO647508
VERSION CO647508.1 GI:41683480
KEYWORDS Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Telford,J., Massignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 4465 02-MAY-2002;
Chiron S.p.A. (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
Location/Qualifiers
1..1731
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTATCTTCC 21
|||||
Db 377 TGCATGTATGGGTATCTTCC 357
|||||
RESULT 27
CO647510/c
LOCUS CO647510 1731 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 4467 from Patent WO234771.
ACCESSION CO647510
VERSION CO647510.1 GI:41683481
KEYWORDS Streptococcus pyogenes
SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Telford,J., Massignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 4467 02-MAY-2002;
Chiron S.p.A. (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
Location/Qualifiers
1..1731
/organism="Streptococcus pyogenes"
/mol_type="unassigned DNA"
/db_xref="taxon:1314"
ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTATCTTCC 21
|||||
Db 377 TGCATGTATGGGTATCTTCC 357
|||||
RESULT 29
AR353656/c
LOCUS AR353656 1803 bp DNA linear PAT 17-AUG-2003
DEFINITION Sequence 5 from patent US 6593093.
ACCESSION AR353656
VERSION AR353656.1 GI:33759689
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE
AUTHORS Uhl,J.R. and Cockerill,F.R.
TITLE Detection of group A Streptococcus
```

JOURNAL	Patent: US 6593093-A 5 15-JUL-2003;	TITLE	Direct Submission
FEATURES	Mayo Foundation for Medical Education and Research; Rochester, MN	JOURNAL	Submitted (24-MAY-1999) Narito Aenuma, Meiji University, Department of Agriculture; Higashimita, Tama-ku, Kawasaki, Kanagawa 214-8571, Japan (E-mail: asanuma@sc.meiji.ac.jp, Tel:+81-44-934-7825(ex.7825), Fax:+81-44-934-7825)
source	1. .1803	COMMENT	On Jan 20, 2000 this sequence version replaced gi:5706360.
ORIGIN	/organism="unknown"	FEATURES	Location/Qualifiers
	/mol_type="genomic DNA"	source	1. .3797
Query Match	100.0%; Score 21; DB 6; Length 1803;		/organism="Streptococcus bovis"
Best Local Similarity	100.0%; Pred. No. 10;		/mol_type="genomic DNA"
Matches	21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		/strain="JB1"
			/db_xref="taxon:1315"
QY	1 TGCATGTATGGTTATCTTCC 21	gene	1775..2042
			/gene="ptsH"
Db	377 TGCATGTATGGTTATCTTCC 357	CDS	1779..2042
			/gene="ptsH"
			/codon_start=1
RESULT 30			/transl_table=1
AX817892/c			/product="histidine containing protein"
LOCUS	AX817892 1803 bp DNA linear PAT 10-DEC-2003		/protein_id="BAA78048.1"
DEFINITION	Sequence 5 from Patent EP1338656.		/db_xref="GI:4958915"
ACCESSION	AX817892		/translation="MASKDPHIVAFGTGHARPATLLVQTASKFASDITLDYKGAVNIL
VERSION	AX817892.1 GI:39723079		KSIMGVMSLGVGGQADVTISAGADADDALAAIETMTKEGLA"
KEYWORDS		gene	2047..3780
SOURCE	Streptococcus sp. 'group A'		/gene="ptsI"
ORGANISM	Streptococcus sp. 'group A',	CDS	2047..3780
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		/gene="ptsI"
	Streptococcus.		/codon_start=1
REFERENCE	1		/transl_table=1
AUTHORS	Uhl, J.R. and cockerill Iii, P.R.		/product="phosphoenolpyruvate-protein phosphotransferase"
TITLE	Detection of group A Streptococcus		/protein_id="BAA78049.1"
JOURNAL	Patent: EP 138656-A 5 27-AUG-2003;		/db_xref="GI:4958916"
FEATURES	MAYO FOUNDATION FOR MEDICAL EDUCATION AND RESEARCH (US)		/translation="MTEMLKGIAASDGVAVAKAYLLVQPDLSFETVTVEDTSAEARL
source	Location/Qualifiers		DAALKASQDELSIIREKAVETIGEEAAVFDHLMVLADPEMISQIKETIRAKQTNAE
	1. .1803		AGLKEVTFMFTIEGEMEDNPMQERAAIDRVAKRVLAHLIGAKLPNPATIDEEISIV
	/organism="Streptococcus sp. 'group A'"		IAHDLTPSDTAQLNKFVKAFVWNGIGTSHSAIMARTLEIAAVLTGNDITSRVKQGD
	/mol_type="unassigned DNA"		IVAVNGITGEVIINPTDSQAEKAEAGVAKQAEWALLKDAKTVTADGKHFELEAAN
	/db_xref="taxon:36470"		IGTPKDVGVNANGAEAVGLYRTFLYMDSDQFPEDQYENAYKAVLSEMGKPVVVR
	/note="ptsI sequence from Oklahoma University M1 strain"		TMDIGGDKELPDLPKEMNPFGLFRALRISISETGNAMFRTIIRALLRASAQVQGLRI
ORIGIN		ORIGIN	MFMVALLKFEAAKAFDEBKANLKAEGVAVSDDIQVIGIMIEIPAAALADQFAKEV
Query Match	100.0%; Score 21; DB 6; Length 1803;		DFFSIGTNDLIQYTMADRMNEQVSYLYQPNPSILRLINNVIKAAHAEKQWGMCGE
Best Local Similarity	100.0%; Pred. No. 10;		MAGDQKAVPLLEVGLDFEFSMSATSLRTRSLMKKLDATKMQEYANRALTEGSTMEEV
Matches	21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		LELSKEYNVND"
QY	1 TGCATGTATGGTTATCTTCC 21	QY	1 TGCATGTATGGTTATCTTCC 21
Db	377 TGCATGTATGGTTATCTTCC 357	Db	2423 TGCATGTATGGTTATCTTCC 2403
RESULT 31		RESULT 32	
AB027569/c		AB006575	
LOCUS	AB027569 3797 bp DNA linear BCT 23-JUL-2003	LOCUS	AB006575 AE004092
DEFINITION	Streptococcus bovis ptsH, ptsI genes for histidine containing protein, phosphoenolpyruvate-protein phosphotransferase, complete cds.	DEFINITION	Streptococcus pyogenes M1 GAS, section 104 of the complete genome.
ACCESSION	AB027569.3 GI:6723275	ACCESSION	AE006575.1 GI:13622480
VERSION	phosphoenolpyruvate-protein phosphotransferase; histidine containing protein.	VERSION	AE006575.1
KEYWORDS	Streptococcus bovis	KEYWORDS	Streptococcus pyogenes M1 GAS
SOURCE	Streptococcus bovis	SOURCE	Streptococcus pyogenes M1 GAS
ORGANISM	Streptococcus bovis	ORGANISM	Streptococcus pyogenes M1 GAS
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
	Streptococcus.		Streptococcus.
REFERENCE	1	REFERENCE	1 (bases 1 to 10430)
AUTHORS	Asanuma, N. and Hino, T.	AUTHORS	Ferretti, J.J., McShan, W.M., Adjic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S.S., Surorov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
TITLE	Molecular characterization of Hpr and related enzymes, and regulation of Hpr phosphorylation in the ruminal bacterium Streptococcus bovis		Complete genome sequence of an M1 strain of Streptococcus pyogenes
JOURNAL	Arch. Microbiol. 179 (3), 205-213 (2003)		
PUBMED	12610726		
REFERENCE	2 (bases 1 to 3797)		
AUTHORS	Asanuma, N. and Hino, T.		

PUBMED REFERENCE AUTHORS	11296296 2 (bases 1 to 10430) Ferrelli, J., McShan, W.M., Adic, D., Savic, D., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Surov, A.N., Kenton, S., Lai, H., Lin, S., Qian, Y., Jia, H.G., Najjar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E. Direct Submission	gene	1AHDLTPSDTAQLNKKQFVKAFVTNIGRTSHSHAINMARTLEIAAVALGTNDITKRVKGDG VIAVNGITGEVIIDPSEDOVLAFKEAGAAAYAKAEASLLKDAHTTETDGGKHFELAA IGTPKDGVDNGDAGVGLYRTEFLYMDSDQPTFEDQYEAAYKAVLBMNGKPVVVR TMDIGDELPEYFDLPKEMNPPLGFRALRISISSETGDMFTOMALLRASVHGOLRI MPPVALLKEFRAAKAVDEKANKLAAGVADDIQVIGIMEIAPAAAMLADQFAKEV DFFSIGTNDLIQYTWADRMNEQVSLYQPTNPISLRLINNVKAAHAEKGKAGCGE MAGDOQAVPLLVGMGLDFMSATSVLRTSLMKKLDLSAKMEEYANRALTECSTAEEV LELSKEYVSED"
TITLE JOURNAL	Submitted (10-APR-2001) Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, 940 SL Young Blvd, Oklahoma City, OK 73104, USA Location/Qualifiers	CDS	complement (3428..3691) /gene="ptsh" /locus_tag="SPy1373" complement (3428..3691) /gene="ptsh" /locus_tag="SPy1373" /note="The N-terminal amino acid sequence of this ORF has been determined from a spot isolated by 2-D gel electrophoresis from another strain of <i>S. pyogenes</i> . Hogan, D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen. 2000. Two-dimensional gel electrophoresis map of <i>Streptococcus pyogenes</i> proteins. Unpublished data.; Best Blastp hit = sp P34366 P7HP STRSL PHOSPHOCARRIER PROTEIN HPR [HISTIDINE-CONTAINING PROTEIN] >gi 480991 pit  837585 phosphotransferase system phosphohistidine-containing protein - <i>Streptococcus salivarius</i> >gi 406779 emb CAA78923.1  (Z17217) HPR [Streptococcus salivarius] >gi 739971 ptf  2004268A His containing phosphocARRIER protein [Streptococcus salivarius] >gi 1583333 prf  2120381A ptsh gene [Streptococcus salivarius]" /codon_start=1 /transl_table=11 /product="putative phosphotransferase system phosphohistidine-containing protein" /protein_id="AAK34200.1" /db_xref="GI:13622483" /translation="MAKDFHVAETGIHARPATLLVQTASKFASDITLDYKGAENVL KSIQVMSLGVGGADVTISAEADAEDAINAIEETMTKEGLA" 4084..4302 /gene="nrhD" /locus_tag="SPy1374" 4084..4302 /gene="nrhD" /locus_tag="SPy1374" /notes="Best Blastp hit = sp Q48708 NRDH LACLC GLUTAREDOXIN-LIKE PROTEIN NRDH >gi 1222680 emb CAA63372.1  (X92690) glutaredoxin-like protein [Lactococcus lactis]" /codon_start=1 /transl_table=11 /product="putative glutaredoxin" /protein_id="AAK34201.1" /db_xref="GI:13622484" /translation="MTIVYKNNCKMCKKFKLEQHGYNFQEIINDEHPEKVDYVKS LGFTSAPVIEADNLVPSGFQPAKLKELI" 4322..6481 /gene="nrhE.2" /locus_tag="SPy1375" 4322..6481 /gene="nrhE.2" /locus_tag="SPy1375" /EC_number="1.17.4.1" /note="Best Blastp hit = gb AAD41036.1 AF112535_3 (AF112535) ribonucleotide reductase alpha-chain [Corynebacterium glutamicum]" /codon_start=1 /transl_table=11 /product="putative ribonucleotide reductase alpha-chain" /protein_id="AAK34202.1" /db_xref="GI:13622485" /translation="MSLKGIDISYERLNNEINRPNVNGKIPLHKDKALKAFSAENVL PNTWSTFSTKELEYLISNDYIESAFIQKRPPIFELDSIKSENFKSPMAVKF YQYVALTKDGEHYLENLDRVFNALYFADGGEDLADLAIVEMINORYOPATPSFLN AGSRREGVSCFLIQVTDMSIGHSINSALQLSRIGGVTLSNLRBAGAPIKGY AGAAAGVVPVWKLFEDFSYSNQLGQRGAGVGVVYNVPHPDIIAFLSTRKENADEKVR
FEATURES source	1. 10430 /organism="Streptococcus pyogenes M1 GAS" /mol_type="genomic DNA" /strain="SF370" /serotype="M1" /db_xref="taxon:160490" complement (78..1505) /gene="gapN" /locus_tag="SPy1371" complement (78..1505) /gene="gapN" /locus_tag="SPy1371" /EC_number="1.2.1.9" /note="The N-terminal amino acid sequence of this ORF has been determined from a spot isolated by 2-D gel electrophoresis from another strain of <i>S. pyogenes</i> . Hogan, D. A., Whitton, M. M., Rogers, J. and R. A. VanBogelen. 2000. Two-dimensional gel electrophoresis map of <i>Streptococcus pyogenes</i> proteins. Unpublished data.; Best Blastp hit = dbj BA05956.1  (AP001514) NADP-dependent glyceraldehyde-3-phosphate dehydrogenase [Bacillus halodurans]" /codon_start=1 /transl_table=11 /product="putative NADP-dependent glyceraldehyde-3-phosphate dehydrogenase" /protein_id="AAK34198.1" /db_xref="GI:13622481" /translations="MAKQYKLVNGEMKLSGENITIVAPATGEELSGVPAMTQAEVDA VYASAKALSDMALSYVERAAYLHKAADILVRDAEKIGAILSKVEAKGHKAASVVI RTAEITNAAEGLMEGEVLEGGSPFAASKKIIAIVRRPGLVLAISPPNYPNVLA GSKIAPALAGNVVAKLPPTQSGISGLLAEAPAEAGIPAGVENTITGRGSVIGDVIV EHAVSFINTGPTTPEGIGIKLAGRNPIMLELGGKSAIVLEDAALAAKNIVAGA FGVSGORCTAKVVLVNDKQVADQLAAEIKTLVLEKLSVGMFEDDADITPLDITSAADFV EGLIKATDKGALTAFNRNGLISPLFDHVTITDMRLAWEEPFGVPLPIIRVTVE EALKISSESYGLQASITFTNFPKAFGIAEQLEVGTVHLNNKTQRTGTDNPFPLGAKKS GAGVQGVKYSIEAMTVKSVVFDIQ" complement (1690..3423) /gene="pstI" /locus_tag="SPy1372" complement (1690..3423) /gene="pstI" /locus_tag="SPy1372" /EC_number="2.7.3.9" /note="Best Blastp hit = sp P30299 P7I_STRSL PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I) >gi 322087 pir  JCI1375 phosphotransferase system enzyme I (EC 2.7.3.9) - <i>Streptococcus salivarius</i> >gi 153615 gb AAK26873.1  (M81756) phosphoenolpyruvate:sugar phosphotransferase system enzyme I [Streptococcus salivarius] >gi 1583334 prf  2120381B ptsh gene [Streptococcus salivarius]" /codon_start=1 /transl_table=11 /product="putative phosphoenolpyruvate:sugar phosphotransferase system enzyme I" /protein_id="AAK34199.1" /db_xref="GI:13622482" /translation="MTMLKGIASDGVAVAKAYLLVQPDLSFETVTVADTNAEEARL DVALQAAQDELSVIRENAVSLGEAAAVFDALHMLVADPEMISQVKETIRAKQTNAE TGLKEVDMFTITEGHNEDPNYMQERAADIDVAKRVLAHLGLVKLNPATINEESIV	gene	complement (1690..3423) /gene="pstI" /locus_tag="SPy1372" complement (1690..3423) /gene="pstI" /locus_tag="SPy1372" /EC_number="2.7.3.9" /note="Best Blastp hit = sp P30299 P7I_STRSL PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE (PHOSPHOTRANSFERASE SYSTEM, ENZYME I) >gi 322087 pir  JCI1375 phosphotransferase system enzyme I (EC 2.7.3.9) - <i>Streptococcus salivarius</i> >gi 153615 gb AAK26873.1  (M81756) phosphoenolpyruvate:sugar phosphotransferase system enzyme I [Streptococcus salivarius] >gi 1583334 prf  2120381B ptsh gene [Streptococcus salivarius]" /codon_start=1 /transl_table=11 /product="putative phosphoenolpyruvate:sugar phosphotransferase system enzyme I" /protein_id="AAK34199.1" /db_xref="GI:13622482" /translation="MTMLKGIASDGVAVAKAYLLVQPDLSFETVTVADTNAEEARL DVALQAAQDELSVIRENAVSLGEAAAVFDALHMLVADPEMISQVKETIRAKQTNAE TGLKEVDMFTITEGHNEDPNYMQERAADIDVAKRVLAHLGLVKLNPATINEESIV

VKTLSLGIITVPDKFYELARKNEDMYLFSPYNVEKEYGIPPNYLDITNNMYDELVANPKI  
TKTKIKARLETEISKLOQESVPYIINIDTANKANPIDGKILMNSCLSEILQVOTPS  
LINDQAEFVEMGTDISCNLGSYNILNMWSPDFGRSIKTKWTALFTVDSSIEAVPT  
IKHNSQAFETGLGANGSYLAQHIEYGPSEIEFTDIYFMLNLTWLTVESNNIAR  
ERQTTTFPENSKIANGSYFDKYVTHGFVPKSDLVKDLFKDHFIFQASDWEALDAVQ  
KDGLTHQNLAVAPNGISYINDCSASIHPIQRIEERQEKIKIYYPANGLSDTDI  
PYVTSAIDMDMRKVIDVYAAATEHVDOGLSLTLFLRSELPMELYEWTKSQKTTTRDLS  
ILRNAYAFNGIKSIYIIRTFTDDGEVGANQCESCVI"  
6814. .7773  
/gene="nrdf.2"  
/locus\_tag="SPy1378"  
6814. .7773  
/gene="nrdf.2"  
/locus\_tag="SPy1378"  
/note="Best Blastp hit = emb|CAA70766.1| (Y09572)  
ribonucleotide reductase subunit R2P [Corynebacterium  
ammoniaenes]"  
/codon\_start=1  
/transl\_table=11  
/product="ribonucleotide diphosphate reductase small  
subunit"  
/protein\_id="AAK34203.1"  
/db\_xref="GI:13622486"  
/translation="MTYYEAINWNEIEDVIDKSTWEKLTQEWLDTRIPLNSDLDDW  
RKLSQEKDLVGVFGGLTLLDTMQSETGVEAIRADVRTPHEEVALNNIQFMSVHAK  
SYSSIFSTLNTKKEIIEIFEWNNHEFLQEKARLINDIYANGDALQKKVASTYLETFL  
FVSGGFTPLYIGNKLANVAEIIKLIIRDESVDHYGYIKFQGLFENELPEDEQENFR  
DMYDILLYQLENEEKYTKLYDGVGWTEEVMTFLRYNANKALMNLGQDPLFPDPTAND  
VNPIVMNGISTGTSNHDFFSQVGVNGYLLGSVEAMSDDDNYGL"  
7748. .9061  
/locus\_tag="SPy1379"  
7748. .9061  
/locus\_tag="SPy1379"  
/note="Best Blastp hit = gb|AAF41386.1| (AE002449)  
chloride channel protein-related protein [Neisseria  
meningitidis Mc58]"  
/codon\_start=1

Query Match 100.0%; Score 21; DB 1; Length 10430;  
Best Local Similarity 100.0%; Pred. No. 7.7;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
TGCATGTATGGGTATCTTCC 3067

Db 3047 TGCATGTATGGGTATCTTCC 3067

RESULT 33  
AE010058  
LOCUS  
DEFINITION  
Streptococcus pyogenes strain MGAS8232, section 106 of 173 of the  
complete genome.  
ACCESSION  
VERSION  
AE010058 AE009949  
KEYWORDS  
SOURCE  
ORGANISM  
Streptococcus pyogenes MGAS8232  
Streptococcus pyogenes MGAS8232  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.  
REFERENCE  
1 (bases 1 to 11992)  
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,  
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,  
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,  
Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Mueser,J.M.  
Genome sequence and comparative microarray analysis of serotype M18  
group A Streptococcus strains associated with acute rheumatic fever  
outbreaks  
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)  
11971708

REFERENCE  
2 (bases 1 to 11992)  
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,  
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,  
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,

TITLE  
Group A Streptococcus strains associated with acute rheumatic fever  
outbreaks

JOURNAL  
PUBMED  
11971708

REFERENCE  
AUTHORS  
Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M.,  
Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M.,  
Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M.,

Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.George. and Mueser,J.M.  
Direct Submission  
Submitted (31-JAN-2002) Laboratory of Human Bacterial  
Pathogenesis/Rocky Mountain Laboratories/NIAID/NIH, 903 S. 4th St.,  
Hamilton, MT 59840, USA  
Location/Qualifiers  
1. .11992  
/organism="Streptococcus pyogenes MGAS8232"  
/mol\_type="genomic DNA"  
/strain="MGAS8232"  
/db\_xref="taxon:186103"  
132. .1073  
/gene="bira"  
132. .1073  
/note="synonym: spyM18\_1374"  
/gene="bira"  
/note="Best blastp match gb|AAK34189.1| (AE006574)  
putative biotin operon repressor [Streptococcus pyogenes  
M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative biotin operon repressor"  
/protein\_id="AAL97969.1"  
/db\_xref="GI:19748528"  
/translation="MKTSEKIVQLLSQTDQDFVSGEYLADQLSISRTSVWKSLSKLENO  
GIQDSLAKHGYVMVQGDILLPKRTISQGLGMPVTYTPHSQSTQDLDAKOGIEAHNSAPR  
LYLAPSGAARGRLDRQFPFSASTGGIYNSMYLKENVPYADMPPTVMVASSIVKAISR  
LTGDTETIKWNDIYLGHNKAVAGILTEAITSVETGLITDVIIVGVGNFVDFDPAIRA  
QKASLFTKEPTITRNLDLIIDIKWLFSLIPVKDHVKYKSEKSLVNLKQVTFIENSQOEK  
RAIADLTDQGLHIVQFDNGDLQTLRSGEISLSWEN"  
complement (1048. .1320)  
/gene="spyM18\_1375"  
complement (1048. .1320)  
/gene="spyM18\_1375"  
/note="Best blastp match gb|AAK34190.1| (AE006574)  
hypothetical protein [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAL97970.1"  
/db\_xref="GI:19748529"  
/translation="MPLTFYLVTFHSTVFFCYNKAMTIRQFLFMAFVCAFETYFFN  
DLLSGNYLFALFWGLLFRDLRRVHTINQLTKTKILKAANSPPKKD"  
complement (1346. .3016)  
/gene="dnaX"  
/note="synonym: spyM18\_1376"  
complement (1346. .3016)  
/gene="dnaX"  
/note="Best blastp match gb|AAF98348.1| (AF280764) DNA  
polymerase III gamma/tau subunits [Streptococcus  
pyogenes]"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma/tau subunits"  
/protein\_id="AAL97971.1"  
/db\_xref="GI:19748530"  
/translation="MYQALRYKRSQTFEDVMVQSVISTTLKQAVESGKISHAYLPSPG  
PRGKTSAAKIFAKMCPNQVGEPCNQCDICRDTNGSLSEDIEDVIEDIAQNSGVDE  
IRDRDKSTYAPSRATYKYVIIDEVHMLSTGAFNALLKTLEETENVVFIATTELHK  
IPATILSRVQRFPAIKQKAIKEITGSISILALGVYRVYVQSOATQALAALETIYDSGK  
ILDQALSLPDNQVTIAEIEITGSISILALGVYRVYVQSOATQALAALETIYDSGK  
SMSRPATLLYLRDLVVKAGDNQROSAVFDTNLSLIDRI FOMITVTVTSHLPEIK  
KGTHTPYAEMMTQLAQEKILSQANLSGELISEIETLKNELAQKQOLQSLQSRPD  
SLAKSDKTKPTTSYRDRVTILKIMEETVRSQSRQYLDLAKNAWEILNDISAQD  
RALMGSPFLANSENALFAEAFNAEQVMNRNNDMFNGIMSKAGFSNLAAP  
RTDFQHIRKEFAQMKSQKDSVQBEQVALDIPEGDFDILLDKINTIDD"  
complement (3016. .3513)  
/gene="spyM18\_1377"  
complement (3016. .3513)  
/gene="spyM18\_1377"  
/note="Best blastp match gb|AAK34192.1| (AE006574)  
conserved hypothetical protein [Streptococcus pyogenes M1  
GAS]"



gene  
CDS  
gene  
CDS  
gene  
CDS  
gene  
CDS

```
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL97972.1"
/db_xref="GI:19748531"
/translation="MNKSKKIQQYOLMIAQAKELFANESNALANI.SNASALINMTLPN
SVTFGLFDGQELILGPFQGRVSCVHKILGKGVCGESQASRRRTIIINDVKQHANYIS
CDAAAMSEIVPMVKEGLIGVLDLSSILVADYDEVQVLEYLTFVDLLLEKTFTTFNM
FGVN"
/genes=".4467
/genes="spym18_1378"
/genes=".4467
/genes="spym18_1378"
/genes="spym18_1378"
/notes="best blastp match gb|AAK34193.1| (AE006574)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL97973.1"
/db_xref="GI:19748532"
/translations="MKQITQQKWLRYGLFIALVINGIELELLGLTANNLSFKKEAFALI
LTISLGIYLIPFAAIFYLKSKFHNLNVIIVSCUGLYISGFLASCGNHLVGQFWS
YIIPSKDALMGDLTAPIVEEPIKASSAILVITLFPRLTKELKVALLSGMQFOL
TEDIRYLOAKSDLSVPTAIGRISTAVTSHVHTAIFTIGAYLLKGNLFSKQQOI
FWLLSPVLGHFIWNSPLTSPGWTLLGLTLLILFGDLFQKINTLDDVLF"
complement(4520..4783)
/genes="spym18_1379"
complement(4520..4783)
/genes="spym18_1379"
/genes="spym18_1379"
/notes="best blastp match gb|AAK34194.1| (AE006574)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL97974.1"
/db_xref="GI:19748533"
/translations="MLLIWGLLFVNNELVFVIYAFDKRKAIKKKRISRNLVITV
LFGFGALLAAKXYHKTRKWFVITWTSILLTLVTFYFMK"
complement(4863..5489)
/genes="udk"
/notes="synonym: spym18_1380"
complement(4863..5489)
/genes="udk"
/notes="best blastp match gb|AAK34195.1| (AE006574)
putative uridine kinase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative uridine kinase"
/protein_id="AAL97975.1"
/db_xref="GI:19748534"
/translations="MFKKPIIIGVTGGGGKTSVSRALDSPFNARIAMIQHDSYK
DQSHMSFEERVKTYNDHPLAFDPMIQOLKELLAGRPVDPIYDYKQHTRSNTFRQ
DPQDVIVTEGLVLEDRLLMDIKLFVDTDDDIRIRIKRDMMERGRSLSEIIDQ
YTSVVKPMYHQFTPEPSKRYADIVPEGVSNVAIDVINSKIASILGEV"
5587..6672
/genes="spym18_1381"
5587..6672
/genes="spym18_1381"
/notes="best blastp match gb|AAK34196.1| (AE006574)
putative RNA helicase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative RNA helicase"
/protein_id="AAL97976.1"
/db_xref="GI:19748535"
/translations="WITKFPPOWEKLDQVAFHLPITQEQAFQPIVDGKNFLGISPT
GTGKTLAYVPSLLALTPKKSQQLILAPNTELAGOIFEVTKWAQPLGLTAQLFISG
TSQRQRLERKQGEIILIGTPRIFELIKLKKIKMGSVNTIVLDEYELLGDSQYDV
QKISHYVPRQHWYMATNKVDQTSIAPNTFCIDLSEQNDIAIQHLYLMVKKRRTD
LLRKFAPIPHFRLAVFNLSLDLGAETQYNGAAVASLADINVKFRKTIIEKPKS
HQLSLLATDLVAGIDIDNLDYVHFDPVARDKENYTHRAGTRMGSKGIVITFVSH
PEDIKKLLKPKAKVSEISLKNQQLHFIN"
complement(6783..8057)
```

CDS  
gene  
CDS  
Query Match  
Best Local Similarity  
Matches  
QY  
Db  
RESULT 34  
AE014229/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

```
/gene="dead2"
/notes="synonym: spym18_1382"
complement(6783..8057)
/genes="dead2"
/notes="best blastp match gb|AAK34197.1| (AE006574)
putative deacetylase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative deacetylase"
/protein_id="AAL97977.1"
/db_xref="GI:19748536"
/translations="KKKLNVLVGLLSILMLSLAIVFINRWKLNEDSORIVLAEKKKN
TSDLVKAVKHIXKDKDYIYFPIKQADDFVDNLPLVSLYKKKSDRELILVKKPLQ
SHLSRVNTLTISKIVYQKFPFLAKSEKVIKSTYHVDNLKPFQVQKDLVSLGRIRIQ
BEVKKYFNAGFNSDKYKLSNLSLDGFEVKSGLNIFDKKLTIPLTFLDFVINPD
FLANSDDRAAYDNYRTYKEQHPKKLVALTDDQPDPTTTTQVLDIILAKYQKGTFFMIG
SKVYVNNELTKVSDAGIEAHNHTWDLNLTNLSVSEIQHVNMTNQALEKACGKKPR
YURPPIGATNATVQSSSLGTQMLWTVDTRDWNHSTDGTMTNVKNQLQFGGVVLMDHI
HQTITNALPTVMEYLKAGYECVTVSELYAHQ"
complement(8153..9580)
/genes="gapN"
/notes="synonym: spym18_1383"
complement(8153..9580)
/genes="gapN"
/notes="best blastp match gb|AAK34198.1| (AE006575)
putative NADP-dependent glyceraldehyde-3-phosphate
dehydrogenase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative NADP-dependent
Query Match 100.0%; Score 21; DB 1; Length 11992;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTATCTTCC 21
|||||
Db 11122 TGCATGTATGGGTATCTTCC 11142
|||||
Streptococcus agalactiae 2603V/R section 39 of 100 of the complete
genome.
AE014229 AE009948
AE014229.1 GI:22533831
Streptococcus agalactiae 2603V/R
Streptococcus agalactiae 2603V/R
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 21781)
Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
Sarelli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
Mariani,M., Vegni,F., Malone,D., Rinaudo,D., Rappuoli,R.,
Telford,J.L., Kasper,D.L., Grandi,G. and Frazer,C.M.
Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
1220547
2 (bases 1 to 21781)
Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A.,
Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M.,
Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A.,
Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J.,
```

Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,  
Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R.,  
Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.  
Direct Submission

JOURNAL  
Submitted (18-JUL-2002) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA

FEATURES  
source

1. .21781  
/organism="Streptococcus agalactiae 2603V/R"  
/mol\_type="genomic DNA"  
/strain="2603V/R"  
/serotype="v"  
/db\_xref="taxon:208435"  
103. .348  
/gene="SAG0813"  
103. .348  
/gene="SAG0813"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAM99700.1"  
/db\_xref="GI:22533832"  
/translation="MNQINHCFSIMVLGVQMKEFIITGRMVARVMETUVCLQWV  
SNLVDMLDIIEINEIVFCVLVASIAGLEFCQKERE"  
387. .674  
/gene="SAG0814"  
387. .674  
/gene="SAG0814"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM99701.1"  
/db\_xref="GI:22533833"  
/translation="MTYIGVPKEQKALRRGKQLEKINOGKTKLMSQMIQDLVGSTCM  
IYLDGSOQTHPYEILKVDQWQIRRNLRKSEKVEIVRIHDIKELRPVNK"  
719. .934  
/gene="SAG0815"  
719. .934  
/gene="SAG0815"  
/note="identified by match to PFAM protein family HMM  
PF01381"  
/codon\_start=1  
/transl\_table=11  
/product="transcriptional regulator, Cro/CI family"  
/protein\_id="AAM99702.1"  
/db\_xref="GI:22533834"  
/translation="MGLILKRLKELRARDGINQTELAKLTGVSRQTISLIBRGVTP  
SVIIAMKISQVPKESLENVPSLVSMEE"  
934. .1695  
/gene="SAG0816"  
934. .1695  
/gene="SAG0816"  
/note="similar to OMNI:NTL01SPL1088; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="membrane protein, putative"  
/protein\_id="AAM99703.1"  
/db\_xref="GI:22533835"  
/translation="WMKNFIKQSDTTFVRIILKSLIGGFIGAILGSVGNLFIFGQDK  
YLSENIVQYFLWRSIVVITLALFSLIYLQIKQKQKQVFNVDSESEIYRQINLR  
HSYGMFTFVSIVLSTVITFNFKLINIFDPSVTLPIDLSLLFLGLHLIYFLKVV  
RNIRIGKMTVAITLKKNNVLQDAELESNYKRCFDIVMNLGFIPTIYFVLVFI  
SFVQKVEIVAIITTSIHIYILKSLKAARHFR"  
complement (1780. .2343)  
/gene="SAG0817"  
complement (1780. .2343)  
/gene="SAG0817"  
/note="similar to OMNI:NTL01SPL0446; identified by  
sequence similarity; putative"  
/codon\_start=1

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/transl\_table=11  
/product="membrane protein, putative"  
/protein\_id="AAM99704.1"  
/db\_xref="GI:22533836"  
/translation="MSKNNTTCLIIETAIFAALAWALSMIPDFASWFTFSFGAIPLLI  
FALRRGTGYFAGLIGLHFLVSKVYLSQVFIYILAFISMGLAGVFSKAFKD  
ALSSTKTKALSIGAILATLVRYVMHYIAGVIFWASYAPKGMSTATLSLSVNGTA  
GLTLFFVVISIILIVISYPSFELPKK"  
complement (2584. .3543)  
/gene="nrdf-2"  
/note="synonym: SAG0818"  
complement (2584. .3543)  
/gene="nrdf-2"  
/note="identified by match to PFAM protein family HMM  
PF00268"  
/codon\_start=1  
/transl\_table=11  
/product="ribonucleoside-diphosphate reductase 2, beta  
subunit"  
/protein\_id="AAM99705.1"  
/db\_xref="GI:22533837"  
/translation="MTTYIEAINNNEIEDVIDKSTWEKLTQFWLDTRIPLSNLDLDDW  
RKLQAQKDLVKGKVGGLTLLDTWQSETGVEAIRADVRTPHEEAVLNNIQMESVHAK  
SYSGTFLNTKSEIEIEFEWNNNEFLQEKARIINDIYANGNALQKKVASTYLENPL  
YSGPFTPLYYLGNKLANVAEIIKLIIRDESVMGTGYGKQFCNKLPLDDEQENFR  
DMYDLDLVQVLENKEKVTKLYDGVGTEEVMTLRYNANKALNGLQDPLFPDPTAND  
VNPVMMGISTGTSNHHDFPSQVNGYLLGSVEAMHDDDDNYGL"  
complement (3746. .5905)  
/gene="nrdb-2"  
/note="synonym: SAG0819"  
complement (3746. .5905)  
/gene="nrdb-2"  
/note="identified by match to PFAM protein family HMM  
PF00317"  
/codon\_start=1  
/transl\_table=11  
/product="ribonucleoside-diphosphate reductase 2, alpha  
subunit"  
/protein\_id="AAM99706.1"  
/db\_xref="GI:22533838"  
/translation="MSLKNIGDVSFYRLNNEINRPVNGQIPLHKDKREALTAFKENVQ  
PNSKAFSDITKIAYLDYLEAFNLKYRPEFTEELSTKLFOKKFRKSGFMAAYKFK  
YQVALKNDGEYVLESTEDRFLNALYFADGDEBELATDLALEISQRYPATPSFLN  
AGSRPGLVSCFLIOVTDNNAIGRSINSALQLSRIGGCVGISLNLREAGAPIKGF  
AGASGVVPVVKLFEDSEFSYNQLQRCQAGVYLDVHPDIIIFSLSTKKENADEKVR  
VKTLSLGITVPDKFTELARNQEMTLFSPYSIEREYGVFPFSDIDITEKIDELVANPNI  
TKINARDLETEISKLQESGYPIINIDTARTNPVDGKIIMSNLCEILQVQKPS  
LINDAQEYLEMGTDI SCNLGSTNVLNMMTSPDFGKSIKMTTRALTFTVDSNIEAVPT  
IKNGQAQHTFGLGAMGLHSLAKNHIYSGSPESIEFTDIYFMLMNYTLVSNNNIAR  
EROTTFVGFESKYADGTYFDKYSGKFVQSDKVKSLFANHFTPKADKWNLRVAVM  
KDGLYHONRLAVAPNGSISYINDCSASIHPIQRIERQEKKIGKIYYPANGLATDTI  
PYTTSYDMDMRKVIDVYAAATEHVDQGLSMTLFLRSEERQEKIKGIYYPANGLATDTI  
ILRNYAFNKGVKSIYYITFTDDGSEVGANQCESCVI"  
complement (5983. .6207)  
/gene="nrdb"  
/note="synonym: SAG0820"  
complement (5983. .6207)  
/gene="nrdb"  
/codon\_start=1  
/transl\_table=11  
/product="ribonucleoside-diphosphate reductase 2,  
Nrdh-redoxin"  
/protein\_id="AAM99707.1"  
/db\_xref="GI:22533839"  
/translation="MAAIVFESKNCMQCKMTKKFLDQHGADFBFEINIDKPEKIEYY  
KNLGFSAAPVIEAGNVVFSQFQPSKLELV"  
6590. .6853  
/gene="ptsh"  
/note="synonym: SAG0821"  
6590. .6853  
/gene="ptsh"  
/note="identified by match to PFAM protein family HMM  
PF00381"

gene

CDS

gene

CDS

```
/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AAM99708.1"
/db_xref="GI:225333840"
/translation="NASKDFHIVAEHTGHARPATLIVQTASKFASDITLDYKGAVALN
KSMIGVMSLGVGGQDVTISAEGADADDAIAAIEETMTKEGLA"
6858..8591
/genes="ptei"
/notes="synonym: SAG0822"
6858..8591
/genes="ptei"
/notes="identified by match to PFAM protein family HMM
PF00391"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAM99709.1"
/db_xref="GI:225333841"
/translation="WTEMLKGIAASDGVAVAKAYLLVQDLSFETVTVEDTNAEEARL
DVALQSDLSVIREKAVESGEAAAVFDHLVLSDFEMINQIKETIRAKQVNAE
TGLKEVDFMTIFEGEDNPYQERADIRDVAKRVLAHLGLVLPNPATINEESIV
IAHDLTPSDAQNLKQFVAFVNTNIGRTSHSAIMARTLEIAAVALGNDITERVQDQ
Query Match 100.0%; Score 21; DB 1; Length 21781;
Best Local Similarity 100.0%; Pred. No. 6.8; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0;
QY 1 TGCATGTATGGTTATCTTCC 21
Db 7234 TGCATGTATGGTTATCTTCC 7214
;
;
RESULT 35
AX602188/c
LOCUS 29072 bp DNA linear PAT 17-FEB-2003
DEFINITION Sequence 117 from Patent WO2092818.
ACCESSION AX602188
VERSION AX602188.1 GI:28402057
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Glaser, P., Rusniok, C., Chevalier, F., Frangeul, L., Lalioui, L.,
Zouine, M., Couve, E., Buchrieser, C., Poyart, C., Trieu-Cuot, P. and
Kunst, F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 117 21-NOV-2002;
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source 1..29072
location/Qualifiers
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 29072;
Best Local Similarity 100.0%; Pred. No. 6.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCATGTATGGTTATCTTCC 21
Db 12272 TGCATGTATGGTTATCTTCC 12252
;
;
RESULT 36
AE014157
LOCUS 51454 bp DNA linear BCT 19-JUL-2002
DEFINITION Streptococcus pyogenes MGAS315, section 22 of 37 of the complete
```

```
genome.
AE014157 AE014074
AE014157.1 GI:21904768
Streptococcus pyogenes MGAS315
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 51454)
Beres, S.B., Sylva, G.L., Barbian, K.D., Lei, B., Hoff, J.S.,
Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F.,
Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and
Musser, J.M.
Genome sequence of a serotype M3 strain of group A Streptococcus:
Phage-encoded toxins, the high-virulence phenotype, and clone
emergence
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
2 (bases 1 to 51454)
Beres, S.B., Sylva, G.L., Barbian, K.D., Lei, B., Hoff, J.S.,
Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F.,
Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and
Musser, J.M.
Direct Submission
Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
Hamilton, MT 59840, USA
1..51454
Location/Qualifiers
/organism="Streptococcus pyogenes MGAS315"
/mol_type="genomic DNA"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/notes="group: A"
complement(150..1139)
/genes="grb"
/notes="synonym: SpvM3_1032"
complement(150..1139)
/genes="grb"
/notes="best blastp hit: gb|AAD26342.1|AF124403.1
(AE014003) GRAB precursor [Streptococcus pyogenes]"
/codon_start=1
/transl_table=11
/product="protein G-related alpha 2M-binding protein"
/protein_id="AAM79639.1"
/db_xref="GI:21904769"
/translation="MGKEIKVKYFLRRSAFGLVAVASVLGVSTVAVDSPIRQPIII
PNGRITNLIGNAPEKALRNEERAIDELKQAIEDKEATTAEASSDALEALADQDA
DALQSEAAVQSDTAAADALEADQTDALQSEAAVVKADNAASDALEALADQTDALQ
VOSEASVVKADTAAADALEADQTDALQSEAAVVKADTAAADALEALADQTDALQ
SEAAVVKADNAASDALEALADQTDALQSEASVVKADTAAADALEALADQTDALQ
TKDTPVKKERQNVNLTPTTGEESNPFPTAALAMVSTGALVSSCKKEN"
complement(1496..2710)
/genes="murZ"
/notes="synonym: SpvM3_1033"
complement(1496..2710)
/genes="murZ"
/notes="best blastp hit: gb|AAL97966.1| (AE010057) putative
UDP-N-acetylglucosamine 1-carboxyvinyltransferase
[Streptococcus pyogenes MGAS8232]"
/codon_start=1
/transl_table=11
/product="putative UDP-N-acetylglucosamine
1-carboxyvinyltransferase"
/protein_id="AAM79640.1"
/db_xref="GI:21904770"
/gene
CDS
```

gene  
complement(2929..4125)  
/gene="metk"  
/note="synonym: SpyM3\_1034"  
CDS  
complement(2929..4125)  
/gene="metk"  
/note="best blastp hit: gb|AAL97967.1| (AE010057)  
S-adenosylmethionine synthetase [Streptococcus pyogenes MGAS82321]  
/codon\_start=1  
/transl\_table=11  
/product="putative S-adenosylmethionine synthetase"  
/protein\_id="AAM79641.1"  
/db\_xref="GI:21904771"  
/translation="MSERKLFITSESVSEGHDPKIDQISDAILDAILAEDPBAHVAEE  
TCVYGSVHFGEISTAYIDNRVVRDTIAEIGTEAEYFSAESVGVHPHSVQSG  
DIAQGVNEALREGDTDLISHAGDQMGFCFAINTEPELMPPIISLSHQLVRLLA  
ELRQGEISYLPDAKSQVTVYDEHDKPVRVDTVVISQHDPEATNDQIRQDVEKV  
IKAVIPADYLDQDTKFFINPTGRFVITGGPQSGSLTGRKIIIVDTYGVSRHGGAFSG  
KQATVKDRSASVAARYIAKLVNAGLATKAELQVLAIGAIVQAQPVSVRVDTFGTSTVPE  
AVLEAAVRQVDFLRPAGIIOQLDLKRPYIKQTAAYGHMGRITDIDLFWHEHLNVDALPE  
AVKTVL"  
complement(4662..7040)  
/gene="inla"  
/note="synonym: SpyM3\_1035"  
CDS  
complement(4662..7040)  
/gene="inla"  
/note="best blastp hit: gb|AAK34188.1| (AE006574) putative  
internalin A precursor [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative internalin A precursor"  
/protein\_id="AAM79642.1"  
/db\_xref="GI:21904772"  
/translation="MKTKKVILIVGLLSQLTLACQSRGNGTYPIKTKQRKQGMTS  
NKIPTKSKKNTKTHGVAGVDFPTDGDGFIITKDSKILSKTDQGVVDHGDHSHFIF  
YADLKSPPEYLIPKGLAKPAVAQASQGTSKVADPHHYEENPADIVAEDALGY  
TVRHDDHFHYILKSSLSGGTQAKQAVATRLPQTSFVSTATANGIPGLHFTPSDGFQ  
FNGQGVGTQKONILVDHGHLPISFADLRGGWAHVADQYDPKAAKPAETHHTP  
ELSERKEYQEKLAYLAELKIDPSTIKRVEVTDGKLGLEYPHPHDAHVLMLSDEIG  
KDIPOPAHEARELEKHKFGMDTLAALGFDEBVIDIVRTHDAPTPFPSPNEKDPNM  
KWLATVIKLDLSGRKDPQRLQKLSLPLETIGFTPIKDISPVLQPKLKOLLMT  
KTGTDTYRFLDNWQLEGDIIQONLKDIFSLSKYKNLTVAANGIDIESLGOLP  
NKLTVLSNNKISDLSPLASLQQLHLDNNQITDLSFVSHKESLTVDVLSRNDVD  
LUTAPLKLMTVNDTKVSHLDPLFKNNPNLSLSINRAQSLLEGIEASSVIVRVEA  
EGNQIKSLVKDQKSGTLFDVTGNQLTSLGYNFTALDIISVKNQNTYNVLSKPN  
KVTNIDISHNNIPLADLNEQHIPEAIKPNFPAYVEGSMVGNNGTAEEKAMATNAK  
ESAOEASESHDYNHNHAYEDEGHAHEHRDKDDHDEHEDENEAKDEQNHAH"  
7244..8185  
/gene="birA"  
/note="synonym: SpyM3\_1036"  
7244..8185  
/gene="birA"  
/note="best blastp hit: gb|AAK34189.1| (AE006574) putative  
biotin operon repressor [Streptococcus pyogenes M1 GAS]"  
/codon\_start=1  
/transl\_table=11  
/product="putative biotin operon repressor"  
/protein\_id="AAM79643.1"  
/db\_xref="GI:21904773"  
/translation="MKTSEKIYQLLSQTDQDFVSGEYLADQLSISRTSVKMSIKSLENG  
GQIDSLKRGYHMQVDILLPLISQGLMPVYTPPHSQSTQLDKAGQIEAHNSAPR  
LYLAPSQEAAKGRDROFFSASTGIIYMSYLPKNVPYADMPPTNMVFASTVVKALISR  
LTGIDTEIKWNDIYLNHNVKAGILTEAITSVETGLITDVIIGVGLNFVTFPPEAIA  
QKAGSLFTBKPTIRNDLIIDWLKFLSPVKDHWKVKYKESLVLNKLQVTFIENSQOEK  
RAIADILTDQGHILVQFENGDLQTLRSGBISLSWEN"  
complement(8160..8444)  
/gene="SpyM3\_1037"  
complement(8160..8444)  
/gene="SpyM3\_1037"  
/note="best non-GAS blastp hit: gb|AAK74993.1| (AE007392)  
hypothetical protein [Streptococcus pneumoniae TIGR4], and  
gb|AAK99574.1| (AE008452) hypothetical protein

[Streptococcus pneumoniae R6]"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM79644.1"  
/db\_xref="GI:21904774"  
/translation="MGKVLPLPTFYLVTFHFHSTVFVCYNKAMTIRQFLMFVCAFET  
YFFNDLLSGNYLFALFWGLLFLRDLRRVHINQLTKILKANSPPKKD"  
complement(8448..10118)  
/gene="dnax"  
/note="synonym: SpyM3\_1038"  
complement(8448..10118)  
/gene="dnax"  
/note="best blastp hit: gb|AAL97971.1| (AE010058) DNA  
polymerase III gamma/tau subunits [Streptococcus pyogenes MGAS82321]"  
/codon\_start=1  
/transl\_table=11  
/product="DNA polymerase III gamma/tau subunits"  
/protein\_id="AAM79645.1"  
/db\_xref="GI:21904775"  
/translation="MYQALYRKYSQTFDEMVGQSVITTLKQAVESGKISHAYLPSG  
PRGTGKSAKIFAKAMNCPQVDEPCNQCDICRDTNGSLEDVIEIDAASNNQVDE  
IRDIRKSTYAPSRATKYKVIIDEVHMLSTGAFNALKLEETPVNVFIATTLHK  
IPATILSRVQRFPAIKQKAIREHLAWLDKEGIAEYVDALNLIARAEGMRDALS  
ILDOALSLSPNQVNTIAIAREITGSISMLAGDYVYVYVSOQATQALAALETIYDSGK  
SMRSPATDILLYRLDLLVYKAGDNQORSAYPDNLSLSDIRIFQMTVTVSHLSEIK  
KQTHPRIYAEMWTIOAQKEQLSQANLSGELISIEIETIKNELAQKQLSQLSQSRPD  
SLASDKTKPKTASVYRVDRTILKIMEETVRNSQSRQYLQDALKNWNEILDNISQD  
RALLMGSEPLVNASENALFAEAFNAEQVMNRNLMFGNIMSKAAGFSFNILAVP  
RTDFQIRKEFAQQMKSQKDSVOEBQEVALDIPEGDFLLDKINTIDD"  
complement(10118..10681)  
/gene="SpyM3\_1039"  
complement(10118..10681)  
/note="best non-GAS blastp hit: gb|AAK06294.1| (AE006448)\_12  
(AE006448) hypothetical protein [Lactococcus lactis subsp.  
lactis]"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM79646.1"  
/db\_xref="GI:21904776"  
/translation="MTEAKGRKKYSNRPVAILKGMKSKKIEOYQOLMIAQAKELFA  
NESNALANLSNALLNMTLNSVPAGFYLDGQBLILGPFQGRVSCVHILKLGKVCQ  
ESAQSRRTIINDVQKHANYISCDAAAMSEIIVPMVKEGHLIGVLDLSSLVADYDEV  
DQYLEAFVDFLEKTTFTFNMFGYKN"  
10760..11569  
/gene="SpyM3\_1040"  
10760..11569  
/note="best non-GAS blastp hit: gb|AAK98831.1| (AE008387)  
Query Match 100.0%; Score 21; DB 1; Length 51454;  
Best Local Similarity 100.0%; Pred. No. 5.9;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACATGATGGTATCTTCC 21  
|||||  
DB 18222 TGCATGTATGGTATCTTCC 18242  
RESULT 37  
BA000034\_07/c  
WPCOMMENT  
Sequence split into 19 fragments LOCUS BA000034 Accession BA000034  
Fragment Name Begin End  
BA000034\_00 1 110000  
BA000034\_01 100001 210000  
BA000034\_02 200001 310000  
BA000034\_03 300001 410000  
BA000034\_04 400001 510000  
BA000034\_05 500001 610000

BA000034\_06 600001 710000  
BA000034\_07 700001 810000  
BA000034\_08 800001 910000  
BA000034\_09 900001 1010000  
BA000034\_10 1000001 1110000  
BA000034\_11 1100001 1210000  
BA000034\_12 1200001 1310000  
BA000034\_13 1300001 1410000  
BA000034\_14 1400001 1510000  
BA000034\_15 1500001 1610000  
BA000034\_16 1600001 1710000  
BA000034\_17 1700001 1810000  
BA000034\_18 1800001 1894275  
Continuation (8 of 19) of BA000034 from base 700001 (BA000034 Streptococcus pyogenes SS1

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 107826 TGCATGTATGGGTATCTTCC 107806

RESULT 38  
BA000034\_08/c  
WPCOMMENT

Sequence split into 19 fragments LOCUS BA000034 Accession BA000034  
Fragment Name Begin End  
BA000034\_00 1 110000  
BA000034\_01 100001 210000  
BA000034\_02 200001 310000  
BA000034\_03 300001 410000  
BA000034\_04 400001 510000  
BA000034\_05 500001 610000  
BA000034\_06 600001 710000  
BA000034\_07 700001 810000  
BA000034\_08 800001 910000  
BA000034\_09 900001 1010000  
BA000034\_10 1000001 1110000  
BA000034\_11 1100001 1210000  
BA000034\_12 1200001 1310000  
BA000034\_13 1300001 1410000  
BA000034\_14 1400001 1510000  
BA000034\_15 1500001 1610000  
BA000034\_16 1600001 1710000  
BA000034\_17 1700001 1810000  
BA000034\_18 1800001 1894275  
Continuation (9 of 19) of BA000034 from base 800001 (BA000034 Streptococcus pyogenes SS1

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 7826 TGCATGTATGGGTATCTTCC 7806

RESULT 39  
CP000003\_10  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000003 Accession CP000003  
Fragment Name Begin End  
CP000003\_00 1 110000  
CP000003\_01 100001 210000  
CP000003\_02 200001 310000  
CP000003\_03 300001 410000  
CP000003\_04 400001 510000  
CP000003\_05 500001 610000  
CP000003\_06 600001 710000  
CP000003\_07 700001 810000  
CP000003\_08 800001 910000

CP000003\_09 900001 1010000  
CP000003\_10 1000001 1110000  
CP000003\_11 1100001 1210000  
CP000003\_12 1200001 1310000  
CP000003\_13 1300001 1410000  
CP000003\_14 1400001 1510000  
CP000003\_15 1500001 1610000  
CP000003\_16 1600001 1710000  
CP000003\_17 1700001 1810000  
CP000003\_18 1800001 1899877  
Continuation (11 of 19) of CP000003 from base 1000001 (CP000003 Streptococcus pyogenes

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 92981 TGCATGTATGGGTATCTTCC 93001

RESULT 40  
CP000017\_10  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000017 Accession CP000017  
Fragment Name Begin End  
CP000017\_00 1 110000  
CP000017\_01 100001 210000  
CP000017\_02 200001 310000  
CP000017\_03 300001 410000  
CP000017\_04 400001 510000  
CP000017\_05 500001 610000  
CP000017\_06 600001 710000  
CP000017\_07 700001 810000  
CP000017\_08 800001 910000  
CP000017\_09 900001 1010000  
CP000017\_10 1000001 1110000  
CP000017\_11 1100001 1210000  
CP000017\_12 1200001 1310000  
CP000017\_13 1300001 1410000  
CP000017\_14 1400001 1510000  
CP000017\_15 1500001 1610000  
CP000017\_16 1600001 1710000  
CP000017\_17 1700001 1810000  
CP000017\_18 1800001 1838554  
Continuation (11 of 19) of CP000017 from base 1000001 (CP000017 Streptococcus pyogenes

Query Match 100.0%; Score 21; DB 1; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 5.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 96984 TGCATGTATGGGTATCTTCC 97004

RESULT 41  
CP000056\_11  
WPCOMMENT

Sequence split into 19 fragments LOCUS CP000056 Accession CP000056  
Fragment Name Begin End  
CP000056\_00 1 110000  
CP000056\_01 100001 210000  
CP000056\_02 200001 310000  
CP000056\_03 300001 410000  
CP000056\_04 400001 510000  
CP000056\_05 500001 610000  
CP000056\_06 600001 710000  
CP000056\_07 700001 810000  
CP000056\_08 800001 910000  
CP000056\_09 900001 1010000  
CP000056\_10 1000001 1110000  
CP000056\_11 1100001 1210000

CP000056_12	1200001	1310000	Continuation (12 of 19) of CP000056 from base 1100001 (CP000056 Streptococcus pyogenes M								
CP000056_13	1300001	1410000									
CP000056_14	1400001	1510000									
CP000056_15	1500001	1610000									
CP000056_16	1600001	1710000									
CP000056_17	1700001	1810000									
CP000056_18	1800001	1897573									
Query Match 100.0%; Score 21; DB 1; Length 110000;											
Best Local Similarity 100.0%; Pred. No. 5.2;											
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;											
QY	1	TGCATGTATGGGTATCTTCC	21								
DB	20289	TGCATGTATGGGTATCTTCC	20309								
RESULT 42											
SAG766847/C											
LOCUS	SAG766847	167050 bp	DNA	linear	BCT 16-APR-2005						
DEFINITION	Streptococcus agalactiae NEM316 complete genome, segment 5.										
ACCESSION	AL766847	AL732656									
VERSION	AL766847.1	GI:23095200									
KEYWORDS	Streptococcus agalactiae NEM316										
SOURCE	Streptococcus agalactiae NEM316										
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.										
REFERENCE	1										
AUTHORS	Glaser, P., Rueniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Meadek, T., Zouine, M., Couve, E., Lalioui, L., Poyart, C., Trieu-Cuot, P. and Kunst, F.										
TITLE	Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease										
JOURNAL	Mol. Microbiol. 45 (6), 1499-1513 (2002)										
PUBMED	12354221										
REFERENCE	2										
AUTHORS	Glaser, P., Rueniok, C. and Frangeul, L.										
TITLE	Direct Submission										
JOURNAL	Submitted (31-MAY-2002) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86										
FEATURES	Location/Qualifiers										
source	1..167050										
	/organism="Streptococcus agalactiae NEM316"										
	/mol_type="genomic DNA"										
	/strain="NEM316"										
	/serovar="serovar III"										
	/db_xref="taxon:211110"										
	join(288..293,303..1412)										
	/genes="gbs0767"										
	288..293										
	/genes="gbs0767"										
	303..1412										
	/genes="gbs0767"										
	/note="Similar to riboflavin specific deaminase (diaminoxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino)uracil reductase)"										
	/codon_start=1										
	/transl_table=11										
	/product="Unknown"										
	/protein_id="CAD46411.1"										
	/db_xref="GI:23095201"										
	/db_xref="GOA:Q8E660"										
	/db_xref="InterPro:IPR002125"										
	/db_xref="InterPro:IPR002734"										
	/db_xref="InterPro:IPR004794"										
	/db_xref="InterPro:IPR011549"										
	/db_xref="UniProt/TREMBL:Q8E660"										
	/translations="NMKEDYNALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGYH										

gene	1381..1387	KRFGLHAEROAIKNADEDISGTLVVTLEPCCHVKGKOPCTEALIKSGIKKVVVGSLL				
RBS	1381..1387	DNPVLSCKGIALLRKECLNVEVGLIRECDALNERFIHMTYKQPFVYLKYAMTLDG				
gene	1393..2043	KIATKTSKWNISNEHSQVQKLRQKCSAIMVGINTVLADNPLRTCTIPKGEALVRI				
CDS	1393..2043	VCDSQLKPLDSILVKSARKTPTWTATQCDNLAAQQOITLKEMCGRLIKVPRKGRKDLK				
		VLMITLQEGIDSLIEGSSLSLHFSALRAGIVNRLVFIAPKIIIGLKAKTAISGEGL				
		DWNQDFRVKDILSRMDSVVIEGKVEHYVYRN"				
	/gene="gbs0768"					
	1381..1387					
	/gene="gbs0768"					
	1393..2043					
	/gene="gbs0768"					
	1393..2043					
	/note="Similar to riboflavin synthase alpha chain"					
	/codon_start=1					
	/transl_table=11					
	/product="Unknown"					
	/protein_id="CAD46412.1"					
	/db_xref="GI:23095202"					
	/db_xref="InterPro:IPR001783"					
	/db_xref="UniProt/TREMBL:Q8E659"					
	/translation="MFTGIIEMGQVSIRINGIKSQQLSIDAPKLVLPLLRKGDSSVAVN					
	GVCLTVLDKSETAFIADVPESMMRTSLAALRLHLSKVLEALRSDSLGSHFVLGHV					
	DGVGKIEIKOKDIAVRFSIDAPPSIMSYIIEKGSVALDGLSLTVVSTFHSFEVSVI					
	PHTMAQTNLSLKKVGDLNIEVDVLGKYAEKFLAPTNRNHTSSVMDWSFLSENGY"					
	2046..3254					
gene	/gene="gbs0769"					
RBS	2046..2051					
	/gene="gbs0769"					
CDS	2061..3254					
	/gene="gbs0769"					
	3,4-dihydroxy-2-butanone-4-phosphate synthase II/					
	/note="Similar to GTP cyclohydrazase II/					
	/codon_start=1					
	/transl_table=11					
	/product="Unknown"					
	/protein_id="CAD46413.1"					
	/db_xref="GI:23095203"					
	/db_xref="GOA:Q8E658"					
	/db_xref="InterPro:IPR000422"					
	/db_xref="InterPro:IPR000926"					
	/db_xref="UniProt/TREMBL:Q8E658"					
	/translation="MAFSPIKKLQDIKSGKMWLMDDENRENGDLICAAEMVTKES					
	INFMAKFGKGLICLPLSNYYAEKLELAQMASHTNDNHTAFTIISIDHLSSTSGIAED					
	RALTAMVANDSSKAKDPRRPGHLFPLLAKEGGLVLRNGHTEATVDLCRLAGLKCCGL					
	CCEINAEDEGSMWRKDELLAFQAQHDLAITIKQLDYRRQEEGGVVRRIEIQLPQFG					
	HFTAYGSEVVANKHEVALVRGDISSEDVLCRLHSECLTGDVPHSLRCDCEQLANA					
	LQOIEAEGRGVLLYMRQGRGIGLINKLKAYHLQEGDLTLEANALGFEGERDYG					
	SAQLKDLGINSINLLTNPNPKIQQLEAEGICVKNRVPLQVAATAVDLNYLTKKXKM					
	GHLDD"					
gene	3255..3739					
RBS	/gene="gbs0770"					
	3255..3260					
	/gene="gbs0770"					
CDS	3269..3739					
	/gene="gbs0770"					
	/note="Similar to riboflavin synthase complex beta chain"					
	/codon_start=1					
	/transl_table=11					
	/product="Unknown"					
	/protein_id="CAD46414.1"					
	/db_xref="GI:23095204"					
	/db_xref="GOA:Q8E657"					
	/db_xref="InterPro:IPR002180"					
	/db_xref="UniProt/Swiss-Prot:Q8E657"					
	/translation="WTIIIEGOLVANEMKIGIVVSRFNELITSKLLSGAVDGLLRHGV					
	BEDIDIVVPGARFIPYARKMALYKDYDAICLGWIKGSTDHYDVCNVEYTKGIGH					
	LNSQSDIPIHFGVLTNDIEQAIERIAAGTKAGNKGVCALSALEMVNLDKLKER"					
gene	complement(3814..5315)					
	/gene="lyss"					
CDS	complement(3814..5304)					
	/gene="lyss"					

```
/codon_start=1
/transl_table=11
/product="lysyl-tRNA synthetase"
/protein_id="CAD46415.1"
/db_xref="GI:23095205"
/db_xref="GOA:Q8E656"
/db_xref="InterPro:IPR002313"
/db_xref="InterPro:IPR004364"
/db_xref="InterPro:IPR004365"
/db_xref="InterPro:IPR006195"
/db_xref="InterPro:IPR008994"
/db_xref="UniProt/Swiss-Prot:Q8E656"
/translation="MNGHIELNDDQIVREKMAALTEQIDPFGKRFERTATSGQL
NEKYADKSDHEDIDETATIGRLMTKRGKGVGAHIDREGQIQIVRKDSVGE
NVEIFPKADGLGVEGQVMTDMGELSIKATHITLSKALRPLPEKFGHLDITFI
YRKRHLDLSNRDSFDRFVTRSKIISEIRKFMDSNGFVETVFLVINEAGASARFPI
THNQADIDMLVRIATLHLKRLIVGGMERVYEIGRIFRNEGMDATHNPEFTSEIAYQ
AVADYQIIDLTEGIQHVTKTVKGDGPNYQGTIEKINEPKFRVMVDVAKVEITGID
FWKEMLEBAQAQKQNPLEKHFVTGVGHIINAFPEEFVEDTLQPTFFVGHPEVVS
FLAKKNDTPRFTRDELFTMTKEYANAFTELNDPIDLSRPEAQASAKELGDEATG
VDYVVEALEYGMPPGGLGIGIDRLCMLLTDTTIRDVLLFFTMK"
complement(5310..5315)
/gene="lysS"
RBS
/gene="lysS"
5467..6411
/gene="gbs0772"
gene
/gene="gbs0772"
5467..5472
RBS
/gene="gbs0772"
5479..6381
CDS
/gene="gbs0772"
/notes="Similar to unknown proteins"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD46416.1"
/db_xref="GI:23095206"
/db_xref="GOA:Q8E655"
/db_xref="InterPro:IPR005833"
/db_xref="InterPro:IPR005834"
/db_xref="InterPro:IPR006439"
/db_xref="UniProt/TREMBL:Q8E655"
/translators="MITSIVFDVDDTIYDQAPYRIAVEKCFDFDFMSAINQAIYRFR
HYSDIGPRVMAGEWTEYFRFWCKETLLLEFGVEIDEATGIYFQBIYHELENIITM
LDSMRITFLKSKNPMGIITNGPTEHQLKVKKGLGYDYDPKRVIVSQATGFKQPK
EKEIFNLAAQFQMNPSFTLLYVGDSDYNDLWGAFGNGHSMWFNHRGRSLKPGIKPVY
DVALDNFEQLFGAVKLVFDLPDKNKFIFDINDKSNPVLNGLNGLMMAERLLESNWS
VDKVVILLRLTAQKQVLRMKYAR"
6390..6411
/gene="gbs0772"
terminator
complement(6417..7069)
/gene="gbs0773"
gene
complement(6417..7058)
/gene="gbs0773"
/notes="Similar to putative phosphoglycerate mutase"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD46417.1"
/db_xref="GI:23095207"
/db_xref="GOA:Q8E654"
/db_xref="InterPro:IPR001345"
/db_xref="UniProt/TREMBL:Q8E654"
/translation="MKFYLRHGKTQWNLGRFGCANGDSPLLEEAIEELEELGOVLS
SIHPDAYSDDLGRADTVNILDNSCPKEIHYTPQLRWALGTLEGCKIATQWAIY
PRQWATYQNPLOQFDMFGAESLYQTHRVESFSLSKSLYDKVLIVGHGANLTAS
IRSLGQYGLSHYKDKLDNLASLIITETHDFKDFNCLTWNDSKYLREQVWTH"
complement(7064..7069)
/gene="gbs0773"
RBS
complement(7080..7553)
/gene="gbs0774"
gene
complement(7080..7553)
/gene="gbs0774"
CDS
/notes="Similar to transcriptional regulator"
/codon_start=1
```

```
/transl_table=11
Query Match 100.0%; Score 21; DB 1; Length 167050;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
|||||
Db 80038 TGCATGTATGGGTATCTTCC 80018

RESULT 43
LOCUS Q655069 349980 bp DNA linear PAT 02-FEB-2004
DEFINITION Sequence 12026 from Patent WO0234771.
ACCESSION Q655069
VERSION Q655069.1 GI:41687946
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Telford,J., Maignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 12026 02-MAY-2002;
Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1..349980
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
/notes="seq 10967 to big: splitted in
seq 10.967, from 0.000.001 to 0.349.980
seq 12.025, from 0.300.001 to 0.649.980
seq 12.026, from 0.600.001 to 0.949.980
seq 12.027, from 0.900.001 to 1.249.980
seq 12.028, from 1.200.001 to 1.549.980
seq 12.029, from 1.500.001 to 1.849.980
seq 12.030, from 1.800.001 to 2.149.980
seq 12.031, from 2.100.001 to 2.160.266"

ORIGIN
Query Match 100.0%; Score 21; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
|||||
Db 191815 TGCATGTATGGGTATCTTCC 191795

RESULT 44
LOCUS AX954529 349980 bp DNA linear PAT 08-JAN-2004
DEFINITION Sequence 1375 from Patent WO03093306.
ACCESSION AX954529
VERSION AX954529.1 GI:40783902
KEYWORDS Streptococcus pyogenes
SOURCE Streptococcus pyogenes
ORGANISM Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Telford,J., Maignani,V., margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a b
JOURNAL Patent: WO 03093306-A 1375 13-NOV-2003;
Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1..349980
/organism="Streptococcus pyogenes"
```



/mol\_type="unassigned DNA"  
/db\_xref="taxon:1314"  
/note="seq 1373 to long, original length: 2.160.266  
replaced by

seq 1373: from 0.000.001 to 0.349.980  
seq 1374: from 0.300.001 to 0.649.980  
seq 1375: from 0.600.001 to 0.949.980  
seq 1376: from 0.900.001 to 1.249.980  
seq 1377: from 1.200.001 to 1.549.980  
seq 1378: from 1.500.001 to 1.849.980  
seq 1379: from 1.800.001 to 2.149.980  
seq 1380: from 2.100.001 to 2.160.266"

ORIGIN

Query Match 100.0%; Score 21; DB 6; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21

Db 220602 TGCATGTATGGGTATCTTCC 220582

RESULT 45

AX569143/c  
LOCUS AX569143 402 bp DNA linear PAT 29-NOV-2002  
DEFINITION Sequence 2351 from Patent WO02077021.  
ACCESSION AX569143  
VERSION AX569143.1 GI:26002638

KEYWORDS Streptococcus pneumoniae

SOURCE Streptococcus pneumoniae

ORGANISM Streptococcus pneumoniae  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE

1  
AUTHORS Massignani,V., Tettelin,H. and Fraser,C.  
TITLE Streptococcus pneumoniae proteins and nucleic acids  
JOURNAL Patent: WO 02077021-A 2351 03-OCT-2002;  
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)  
FEATURES Location/Qualifiers

source

1. .402

/organism="Streptococcus pneumoniae"

/mol\_type="unassigned DNA"

/db\_xref="taxon:1313"

ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 402;  
Best Local Similarity 95.2%; Pred. No. 87;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21

Db 377 TGCATGTATGGGTATCTTCC 357

RESULT 46

AR480287/c  
LOCUS AR480287 1140 bp DNA linear PAT 14-MAY-2004  
DEFINITION Sequence 1191 from patent US 6699703.  
ACCESSION AR480287

VERSION AR480287.1 GI:47239249

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM

Unclassified.  
REFERENCE 1 (bases 1 to 1140)  
AUTHORS Doucette-Stamm,L., Bush,D., Zeng,Q., Opperman,T. and  
Houseweart,C.E.

TITLE Nucleic acid and amino acid sequences relating to Streptococcus  
pneumoniae for diagnostics and therapeutics  
JOURNAL Patent: US 6699703-A 1191 02-MAR-2004;  
Genome Therapeutics Corporation; Waltham, MA  
FEATURES Location/Qualifiers

source

1. .1140

/organism="unknown"

/mol\_type="genomic DNA"

ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1140;  
Best Local Similarity 95.2%; Pred. No. 73;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21

Db 377 TGCATGTATGGGTATCTTCC 357

RESULT 47

AR587350/c

LOCUS AR587350 1287 bp DNA linear PAT 15-DEC-2004

DEFINITION Sequence 1469 from patent US 6800744.

ACCESSION AR587350

VERSION AR587350.1 GI:56631607

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1287)

AUTHORS Doucette-Stamm,L.A. and Bush,D.

TITLE Nucleic acid and amino acid sequences relating to Streptococcus

pneumoniae for diagnostics and therapeutics

JOURNAL Patent: US 6800744-A 1469 05-OCT-2004;

Genome Therapeutics Corporation; Waltham, MA

FEATURES Location/Qualifiers

source

1. .1287

/organism="unknown"

/mol\_type="genomic DNA"

ORIGIN

Query Match 92.4%; Score 19.4; DB 6; Length 1287;  
Best Local Similarity 95.2%; Pred. No. 72;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21

Db 431 TGCATGTATGGGTATCTTCC 411

RESULT 48

STREIPEPA/c

LOCUS STREIPEPA 2259 bp DNA linear BCT 26-APR-1993

DEFINITION Streptococcus salivarius phosphoenolpyruvate:sugar

phosphotransferase system enzyme I (ptsI) gene, complete cds.

ACCESSION M81756

VERSION M81756.1 GI:153614

KEYWORDS cytoplasmic protein; phosphoenolpyruvate:sugar phosphotransferase

system enzyme I; protein kinase; ptsI gene; sugar transport

SOURCE Streptococcus salivarius

ORGANISM Streptococcus salivarius

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

Streptococcus.

REFERENCE 1 (bases 1 to 2259)

AUTHORS Gagnon,G., Vadeboncoeur,C., Levesque,R.C. and Frenette,M.

TITLE Cloning, sequencing and expression in Escherichia coli of the ptsEI

gene encoding enzyme I of the phosphoenolpyruvate:sugar

phosphotransferase transport system from Streptococcus salivarius

Gene 121 (1), 71-78 (1992)

JOURNAL 1427100

PUBMED

COMMENT Original source text: Streptococcus salivarius (library: ATCC

25975) DNA.

FEATURES Location/Qualifiers

source

1. .2259

/organism="Streptococcus salivarius"

/mol\_type="genomic DNA"

/db\_xref="taxon:1304"

gene	/tissue_lib="ATCC 25975"									
-35_signal	22..2023									
TATA_signal	/gene="ptcI"									
CDS	22..27									
	/gene="ptcI"									
	47..52									
	/gene="ptcI"									
	79..1812									
	/gene="ptcI"									
	/codon_start=1									
	/transl_table=11									
	/product="phosphoenolpyruvate:sugar phosphotransferase									
	system enzyme 1"									
	/protein_id="AAA26873.1"									
	/db_xref="GI:153615"									
	/translation="MTKMLKGIASDGVAVAKAYLLVQPDLSFETVTVDDTSABEARL									
	DRLAASQDELSVIREKAVESLGEAAAFDAHLMLVADPEMTGQIKETIRAKQVNAE									
	AALTEVTDNFIAIFEGMEDNPYMQEAAADIRVTKRVLANLIGKKLPNPATINEESIV									
	VAHDLTPSDTAQLNKYKAFVTNIGRTSHSAIMARTLEIAAVLTGNNITELVKGDD									
	ILAVSGITGEVINPTEQIAEFKAAGEAYAKQAEWALLKOAQTVTAGKGFPLAAN									
	IGTPKDVGVNDNGABAVGLITFEFLYMDSQDPPTDDQYEAAYKAVLEGMGKFPVVR									
	TMDIGGDKELPYFDLPKEMNPFILGRALRISISETGNQMFRTQLRALLRASVHGKURI									
	MFPWVALLTEFRTAKGLIEBKAKLVAEGYVADDDIEVGIMIEIPAAAMLADQFAKEV									
	DFPSIGTNDLIQYTMADRMNEQSVLYQPNPSILRLINNVIKAAHABGKWAGMCGE									
	MAGDOTAVPLVGMGLDFMSATSVLRTSLMKKLDTAQMEYANRALTECSTMEEV									
	LELSKYVNDV"									
terminator	1817..1847									
	/gene="ptcI"									
terminator	/note="putative"									
	1996..2023									
	/gene="ptcI"									
	/note="putative"									
RIGIN										
Query Match	92.4%; Score 19.4; DB 1; Length 2259;									
Best Local Similarity	95.2%; Pred. No. 65;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Qy	1 TGCATGTATGGGTATCTTCC 21									
b	455 TGCATGTATGGGTGTCTTCC 435									
RESULT 49										
0789003/c										
OCUS										
DEFINITION	Sequence 94 from Patent EP1400592.									
ACCESSION	CQ789003									
VERSION	CQ789003.1 GI:45822571									
KEYWORDS	unidentified									
SOURCE	unclassified									
ORGANISM	unclassified sequences.									
REFERENCE	1 Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Barash,S.C.,									
AUTHORS	Fannon,M. and Dougherty,B.A.									
TITLE	Streptococcus pneumoniae polynucleotides and sequences									
JOURNAL	Patent: EP 1400592-A 94 24-MAR-2004;									
FEATURES	HUMAN GENOME SCIENCES, INC. (US)									
source	Location/Qualifiers									
	1..8195									
	/organism="unidentified"									
	/mol_type="unassigned DNA"									
	/db_xref="taxon:32644"									
Query Match	92.4%; Score 19.4; DB 6; Length 8195;									
Best Local Similarity	95.2%; Pred. No. 53;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Qy	1 TGCATGTATGGGTATCTTCC 21									
b	1384 TGCATGTATGGGTGTCTTCC 1364									
RESULT 50										
AR218862/c										
LOCUS										
DEFINITION	Sequence 94 from patent US 6420135.									
ACCESSION	AR218862									
VERSION	AR218862.1 GI:23319796									
KEYWORDS	Unknown.									
SOURCE	Unknown.									
ORGANISM	Unclassified.									
REFERENCE	1 (bases 1 to 8195)									
AUTHORS	Kunsch,C.A., Choi,G.H., Dillon,P.S., Rosen,C.A., Barash,S.C.,									
	Fannon,M.R. and Dougherty,B.A.									
TITLE	Streptococcus pneumoniae polynucleotides and sequences									
JOURNAL	Patent: US 6420135-A 94 16-JUL-2002;									
FEATURES	Human Genome Sciences, Inc.; Rockville, MD;									
source	WOX;									
	Location/Qualifiers									
	1..8195									
	/organism="unknown"									
	/mol_type="genomic DNA"									
ORIGIN										
Query Match	92.4%; Score 19.4; DB 6; Length 8195;									
Best Local Similarity	95.2%; Pred. No. 53;									
Matches	20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
Qy	1 TGCATGTATGGGTATCTTCC 21									
b	1384 TGCATGTATGGGTGTCTTCC 1364									
RESULT 51										
BD003774/c										
LOCUS										
DEFINITION	Polynucleotide of Streptococcus pneumoniae and sequence.									
ACCESSION	BD003774									
VERSION	BD003774.1 GI:18631735									
KEYWORDS	unidentified									
SOURCE	unclassified									
ORGANISM	unclassified.									
REFERENCE	1 (bases 1 to 8195)									
AUTHORS	Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,									
	Fannon,M. and Dougherty,B.A.									
TITLE	Polynucleotide of Streptococcus pneumoniae and sequence									
JOURNAL	Patent: JP 2001501833-A 94 13-FEB-2001;									
COMMENT	HUMAN GENOME SCIENCES INC									
	OS Unidentified									
	PN JP 2001501833-A/94									
	PD 13-FEB-2001									
	PF 30-OCT-1997 JP 1998520718									
	PR 31-OCT-1996 US 60/029960									
	PI CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI									
	STEVEN C BARASH,									
	PI MICHAEL FANNON,BRIAN A DOUGHERTY									
	PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,									
	PC C12N1/21,									
	PC C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC									
	G06F15/40									
	CC Strandedness: Double;									
	CC Topology: Linear;									
	FH Key									
	FT source									
	1..8195									
	Location/Qualifiers									
	/organism='Unidentified'.									
FEATURES	Location/Qualifiers									
source	1..8195									
	/organism="unidentified"									
	/mol_type="genomic DNA"									
	/db_xref="taxon:32644"									
ORIGIN										



SKREGINAEQIVIKIDQGVYVTSBGHDHYHYNGKVPYDAIISBEILMKDPNPKDKED  
IYNEVGKGYIVKDGKYYVYLKAAHADNVRTKEENROKQHSQHREGSTPRNDGAV  
ALARSQRYITDDGYIFNADSIITEDTGDIVVPHGDHYHYI PKNLSASELAAAEFL  
SGRGNLSRTYRQSDNTRTNVTNSVNSPGTTNTNNTNNTNSQASNDIDSL  
LKQLYKLPSRHEVSDGLVDPQAITSRTARGVAVPHGDHYHFIPYSQWSELEERIA  
RIIPLRYRSHNVDPDRPEQSPQPTPEPSPQAPAPNLKIDNSLSVLQVRAKVGEG  
YVFEKISRYVFAKLPSETVKNLESKLQSDSVSHTLTAKKENVAPRQDFYDKAY  
NLTEAHKALFENKGNSDFOALDKLLERLNDSTNKEKLVDLLAFLAPIHTPERLG  
KPNQSYTEDEVRIAQLADKYTSDGYIFDEHDIISDEGDVYVTHMGHSHWIKDS  
LSDEKVAQAQYKTEKGLPPSPADVKANFTGSAAIYTRVKGKRIPLVLPLWY  
EHTVEKNGNLIIPKHDHNYIKFAMFDDHTYKAPNGYTLLEDLFAIKYVHEPDRP  
HNSDNGWNASHEVGLKGDHSEDPKNFKADEEPEETPAEPVQVETEKVEAQLEKA  
EVLIAKYTDSLSKANATETLAGLRNLTJLQIMDNNSIMAEAEKLLALLKGSNPSVSK  
EKIN"

gene  
complement (8049. .9782)  
/gene="SP1176"  
CDS  
complement (8049. .9782)  
/gene="SP1176"

/codon\_start=1  
/transl\_table=11  
/product="phosphoenolpyruvate-protein phosphotransferase"  
/protein\_id="AAK75285.1"  
/db\_xref="GI:14972656"  
/translation="MTMLKGIAGSGVAVAKAYLIVQPLSPETITVEDTNAEEARL  
DAALQASODELSVIREKAVCTLGEAAQVFDALMLVLADPEMISOIKETIRAKKNVAE  
AGLKEVTDMEITITPEGMEDPNYQERAADIRDVTKVLNLLGKLPNPASINEEVIV  
IAHDLTSTQAQDKNFVKAPVTNIGGRTSHSAIMARTLEIAAVLGTNNITEIVKGD  
ILAVNGITSVIINPTDEAAEFKAAGEAYAKQAEWALLKDAQTVDAGKHFEAAN  
IGTPKQVGVNNGAEGAVGLYRTEFLYMSQDFPTEDEQYEAQKAVLGNMGKPVVVR  
TMDIGDKELPYPDMEMNPFILGFRALRISISETGDAMFTQIRALLRASVHGQLRI  
MFMVALLKEFRAAKAVDEDEKLANLAEGVAVADNIQVIMIEIPAAAMLADQFAKEV  
DFPSGNLDLIQYTMADRMNEQVSLYQYNPSSLRLINNVIKAAHAEGKAWGCGE  
MAGDQQAFLVGLVGLGDEFMSATSVLRTSLMKGLDTAKNBEYANRALTECSTMREV  
LELQKEYVNF"

gene  
complement (9788. .10051)  
/gene="SP1177"  
CDS  
complement (9788. .10051)  
/gene="SP1177"  
/codon\_start=1  
/transl\_table=11  
/product="phosphocarrier protein Hpr"  
/protein\_id="AAK75286.1"  
/db\_xref="GI:14972657"

/translation="MASKDFHVVAETGIHARPATLLVQTASKPASDITLEYKGSVNL  
KSLNGVMSLGVGGQADVTTISABGADADDAIAAISETWEKEGLA"

ORIGIN

Query Match 92.4%; Score 19.4; DB 1; Length 10256;  
Best Local Similarity 95.2%; Pred. No. 51;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21  
|||||  
Db 9406 TGCATGTATGGGTTGCTCTCC 9426

RESULT 53  
AE008479  
LOCUS  
DEFINITION Streptococcus pneumoniae R6 section 95 of 184 of the complete genome. BCT 13-SEP-2001

ACCESSION AE008479 AE007317  
VERSION AE008479.1 GI:15458677

KEYWORDS  
SOURCE Streptococcus pneumoniae R6  
ORGANISM Streptococcus pneumoniae R6

Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

REFERENCE 1 (bases 1 to 10320)

AUTHORS  
Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczyk, L., Burgett, S.,  
DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,  
Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., LaGace, R.,  
LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

McAhen, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,  
Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rocky, P.,  
Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,  
Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L.  
and Glass, J.I.  
Genome of the bacterium Streptococcus pneumoniae strain R6  
J. Bacteriol. 183 (19), 5709-5717 (2001)  
11544234

2 (bases 1 to 10320)

Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczyk, L., Burgett, S.,  
DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,  
Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., LaGace, R.,  
LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,  
McAhen, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,  
Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rocky, P.,  
Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,  
Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L.  
and Glass, J.I.

Direct Submission

Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA

FEATURES

source

1. .10320  
/organism="Streptococcus pneumoniae R6"  
/mol\_type="genomic DNA"  
/strain="R6"  
/db\_xref="taxon:171101"

gene

complement (91. .738)  
/gene="spr1056"  
CDS  
complement (91. .738)  
/gene="spr1056"  
/codon\_start=1  
/transl\_table=11  
/label="spr1056"

/product="Hypothetical protein"

/protein\_id="AAK99860.1"

/db\_xref="GI:15458678"

/translation="MTTKANITIEEYIEMSEVDFNEAVNVEFTSDTCQLANSYQSL  
FKFDFKNFSGDLIFTWKPSLVKEGDIIGRRDSQVDNLRIAGNIPLNTRKYSLN  
MNRNGCMGDFPHDFDIYLDHVAKYAEQKNVKEYPKLRALLHONALYPLFSN  
FDDFLKKNYKLTITWQVSKETPFPSDFNMFPKNISEKIIIFERGSKMLNDLSKNYK"

gene

complement (752. .1465)  
/gene="spr1057"  
CDS  
complement (752. .1465)  
/gene="spr1057"  
/codon\_start=1  
/transl\_table=11  
/label="spr1057"

/product="Conserved hypothetical protein"

/protein\_id="AAK99861.1"

/db\_xref="GI:15458679"

/translation="MFYKFLFDLHDTLLDFAEDVALTOLLKEGVADIQAYKDYV  
VPMKALWDLLEKLSKQELVNTFRSLRSHFGQEKDQGSFLAORYQFLAQCGTLLS  
GAHDLDSLIERYDYLAATNGITAIQTGLRAQGLVYFNQVFISSQLQKQDPAFL  
YEKIQOIAFGSKETLMIGDSLTDITQGGNNAGIDTIWYNPHPLENHTQAQPTVEY  
SYODLLDKNILEKITP"

gene

complement (1923. .2081)  
/gene="spr1058"  
CDS  
complement (1923. .2081)  
/gene="spr1058"  
/codon\_start=1  
/transl\_table=11  
/label="spr1058"

/product="Hypothetical protein"

/protein\_id="AAK99862.1"

/db\_xref="GI:15458680"

/translation="MNLKWTDFSDKSIINRVPIVEGAYNAALLOQAGELSVIQTO  
LAELINK"

gene

complement (2151. .2306)  
/gene="spr1059"  
CDS  
complement (2151. .2306)  
/gene="spr1059"  
/codon\_start=1

```
/transl_table=11
/label=spr1059
/product="Hypothetical protein"
/protein_id="AAK99863.1"
/db_xref="GI:15458681"
/translation="NMESIGIGLIVIVSHSHKIAEGVVELISKVAKDVPITYVGGTEGG
GIGT9FD"
repeat_region
complement (2542..2613)
/rpt_family="RUP element"
/rpt_type="dispersed"
/rpt_unit="rup.a"
complement (2725..5292)
/gene="pnpa"
/notes="synonym: spr1060"
CDS
complement (2725..5292)
/gene="pnpa"
/codon_start=1
/transl_table=11
/label=spr1060
/product="Histidine Motif-Containing protein"
/protein_id="AAK99864.1"
/db_xref="GI:15458682"
/translation="MNQIYLRKEERMKINKKYLAGSVATLVLSVCAYELGLHQAOTVK
ENNRVYIDGKATQKTENLTPTDEVSKREGINAQIVIKITDQGYVTHSGDHYHYNG
KYPYDAIISELLMKDPNYQLKDEDIISIKGGYIVKDGKYVYLKDAAHADNVRTK
EENRQKQSHQREGGTRNDGAVALARSQGRYTTDDGYIFNADSIEDTGDATVP
HGDHYIYIPKNELASELAAAKAFVLSGRGLNSRTYRRQNSDNTSRTNVPVSVPNG
TNTNTNSNTSQASQNDISLLKQYIKLPLSQRHVESDGLIFDPAQITSTRANG
VAVPDGHVHFIPYSQLSPLEEKLARIIPLYRSNHWPDSPRQSPQSTPEPSPSP
QAPNPQAPNPIDSKLVKEAVRKVDGYVFEENGVPRIYIPAKDLSAETAAGIDSKL
AKQESLSHKLGAKTLDPSDFRYNKAYDILLARIHDLDDNKGROVDFEALDNLIER
LKDVSSDKVLKVDLILAFIAPLRHPERLKGPNQAITYTDEIQVAKLACKYTTEDGYI
FPDRITSDGDAYVTPMTHSHWIKDLSISEARAAQAAYAKEXGLTPPSDHDQSG
NTEAKGAIYINRWKAAPVLDPMFIPNQYTVFVKGSLIIPHDYHNKINFEKDFE
GLYEAPKSLLEDLATVYVHEPNERHSDNGFNASDHVQRNKGNQADTNQTEPK
NEEKPQTEPEETREKPEQSPKPEPTEPESEPESEPESEPESEPESEPESEPESEPE
KLREADLLGKQNPITIKNAKETLTGLKNLLFGTQDNNTIMAEKLLALLKESK"
complement (5417..7903)
/gene="pnta"
/notes="synonym: spr1061"
CDS
complement (5417..7903)
/gene="pnta"
/codon_start=1
/transl_table=11
/label=spr1061
/product="Pneumococcal histidine triad protein A
precursor"
/protein_id="AAK99865.1"
/db_xref="GI:15458683"
/translation="MQLEISNRKUSMKINKYLVGSAALILVSCVELGLYQARTV
KENRVSYIDGQATQKTENLTPTDEVSKREGINAQIVIKITDQGYVTHSGDHYHYN
KGVYDAIIESELLMDPNYKLUDEDIVNEVGKYVIVKDGKYVYLKDAAHADNVRT
KEEINRQKQSHQREGGTRNDGAVALARSQGRYTTDDGYIFNADSIEDTGDATVP
PHGDHYIYIPKNELASELAAAEFLSGRGLNSRTYRRQNSDNTSRTNVPVSVPNG
GTTNTNSNTSQASQNDISLLKQYIKLPLSQRHVESDGLVDFDPAQITSTRAR
GVAVPDGHVHFIPYSQMSLEERIIARIIPLYRSNHWPDSPRQSPQSTPEPSPSP
QAPNLKLSNSQLVLRKVGEGVFEENGVPRIYIPAKDLSAETAAGIDSKL
ESVSHLTATKANVAPRDQFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLD
ESTNKEKLVDDLAFIAPLRHPERLKGPNQIYEYTEDVRIQALADKYTSDGYIFDE
HDIISDEGDAYVTPHMGSHWIKDLSLSDKEKVAQAQYIIPKDHQHYNIKFAFDVDMPT
GSAAIYNRVKGEKIPLVRLPYMVEHTVEVNGNLIIPKDHQHYNIKFAFDVDMPT
KAPNGVTLDELFIKYVHEPDERPHSDNGNASEHVLGKDHSEDNKNKFADEE
PVETPEPEVQVETKEVQAKEAVLLAKVITDSSKANATETIAGURNLITLIQIM
DNNSINAEAKULALLKGNPSVSKSKIN"
complement (8122..9855)
/gene="ptei"
/notes="synonym: spr1062"
CDS
complement (8122..9855)
/gene="ptei"
/EC_number="2.7.3.9"
/codon_start=1
/transl_table=11
/label=spr1062
```

```
/product="PTS enzyme I"
/protein_id="AAK99866.1"
/db_xref="GI:15458684"
/translation="MTEMLKGIASDGVAVAKAYLLVOPDLSFETITVEDTNAEEARL
DAALQASQESLVIREKAVGTLGEEAAQVFDHLMVLADPDMISQIKETIRAKKNVAE
AGLKEVTDMFITIFEGMEDNPYQRAADIRDVTKRVLANLLGKLPNPASINEEVIV
IAHDITPSDITQOLDKNFYKAFVTNIGRTSHSAIMARTLEIAAVLTGTTNNITIEIKDGD
ILAVNGITGEVILNPTDQAAEFKAAGBAYAKQAEWALLKDAQVTVADGKHFEALAN
IGTPKDGVEGNNGAEVGLYRTEFLYVDSQDPTFEDQYEAQKAVLEGMNGKPVVVR
TWDIGGDXELPYFDMPHENPFLGPRALRISISETGDAMFRTQIALLRASVHGOLRI
MPFVALLKEPFAAKAVFDEERKANLIAGVADNIIQVGMIEIPAAAMLAQDQPAKEV
DFFSTGTNDLIQYTMAADRMEQSYLYQPNPSTLRLINNVIKAAHAEGRWAGWKEV
MAGDQAVPLLVGMGLDEBFSMSATSVLRSLMKKLDTKMEEYANRALTECSTMBEEV
LELOKEYVNF"
complement (9861..10124)
/gene="ptsh"
/notes="synonym: spr1063"
CDS
complement (9861..10124)
/gene="ptsh"
/codon_start=1
/transl_table=11
/label=spr1063
/product="Histidine-containing phosphocarrier protein of
the PTS"
/protein_id="AAK99867.1"
/db_xref="GI:15458685"
/translation="WASKDPHVVAETGTHARPATLLVOTASKFASDITILEYKKGKSVNL
KSIQVMSLGVGGQADVTISAGADDDAIAAISETMEKGLA"
ORIGIN
Query Match 92.4%; Score 19.4; DB 1; Length 10320;
Best Local Similarity 95.2%; Pred. No. 51;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TCATGATATGGTTATCTTC 21
|||||
Db 9479 TGCATGTATGGTTGCTCTTC 9499
RESULT 54
SPNEU1915/c
LOCUS SPNEU1915 20035 bp DNA linear HTG 11-JUL-2001
DEFINITION Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.
ACCESSION AL449937
VERSION AL449937.1 GI:11545162
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Humbert,Y.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peitsch,M.
and Garcia-Bustos,J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
PUBMED 11442348
REFERENCE 2 (bases 1 to 20035)
AUTHORS Dopazo,J., Mendoza,A., Herrero,J., Caldara,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Buell,G., Feger,G., Garcia,E., Peitsch,M. and
Garcia-Bustos,J.F.
Direct Submission
JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
COMMENT * NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
source 1..20035
```

```
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/serotype="19f"
/db_xref="taxon:1313"
/clone="G54"

ORIGIN
Query Match      92.4%; Score 19.4; DB 14; Length 20035;
Best Local Similarity 95.2%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 15401 TGCATGTATGGGTATCTTCC 15381

RESULT 55
AX5711763
LOCUS      AX5711763      349980 bp      DNA      linear      PAT 29-NOV-2002
DEFINITION Sequence 4982 from Patent WO02077021.
ACCESSION  AX5711763
VERSION     AX5711763.1 GI:26003955
KEYWORDS   Streptococcus pneumoniae
SOURCE     Streptococcus pneumoniae
ORGANISM   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE  1 Masignani,V., Tettelin,H. and Fraser,C.
AUTHORS   Masignani,V., Tettelin,H. and Fraser,C.
TITLE     Streptococcus pneumoniae proteins and nucleic acids
JOURNAL   Patent: WO 02077021-A 4982 03-OCT-2002;
Chiron Spa (IT) ; THE INSITUITE FOR GENOMIC RESEARCH (US)
FEATURES   Location/Qualifiers
            source      1..349980
                        /organism="Streptococcus pneumoniae"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:1313"
                        /notes="seq 4979 too long: 2.162.598 bases
                        replaced by following seq:
                        seq 4979: from 0.000.001 to 0.349.980
                        seq 4980: from 0.300.001 to 0.649.980
                        seq 4981: from 0.600.001 to 0.949.980
                        seq 4982: from 0.900.001 to 1.249.980
                        seq 4983: from 1.200.001 to 1.549.980
                        seq 4984: from 1.500.001 to 1.849.980
                        seq 4985: from 1.800.001 to 2.149.980
                        seq 4986: from 2.100.001 to 2.162.598"

ORIGIN
Query Match      92.4%; Score 19.4; DB 6; Length 349980;
Best Local Similarity 95.2%; Pred. No. 28;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 108618 TGCATGTATGGGTATCTTCC 108638

RESULT 56
AP001046
LOCUS      AP001046      158063 bp      DNA      linear      PRI 14-JAN-2000
DEFINITION Homo sapiens genomic DNA, chromosome 21, clone:KB43F12, MX1-D21S171
region, complete sequence.
ACCESSION  AP001046
VERSION     AP001046.1 GI:6693596
KEYWORDS   HTG.
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo
REFERENCE  1 (bases 1 to 158063)
AUTHORS   Shimizu,N., Kudoh,J. and Shibuya,K.

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="21"
/clone="KB43F12"
/cell_line="FLEB14-14"
/cell_type="pre-pro-B cell"
/clone_lib="Keio BAC library"

FEATURES   source      1..158063
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosomes="21"
                        /clone="KB43F12"
                        /cell_line="FLEB14-14"
                        /cell_type="pre-pro-B cell"
                        /clone_lib="Keio BAC library"

ORIGIN
Query Match      87.6%; Score 18.4; DB 8; Length 158063;
Best Local Similarity 95.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 20
    |||||
Db 90265 TGCATGTATGAGTTATCTTCC 90284

RESULT 57
AC144901/c
LOCUS      AC144901      178293 bp      DNA      linear      HTG 22-JUL-2003
DEFINITION Sus scrofa clone RP44-138L19, WORKING DRAFT SEQUENCE.
ACCESSION  AC144901
VERSION     AC144901.3 GI:331112720
KEYWORDS   HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE     Sus scrofa (pig)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
REFERENCE  1 (bases 1 to 178293)
AUTHORS   Antonellis,A., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Cariaga,K.,
Chu,G., Coleman,B., Coleman,H., Engle,J., Granite,S., Guan,X.,
Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Hu,P.,
Hurle,B., Idol,J.R., Karlins,E., Kwong,P., Laric,P., Lee-Lin,S.-Q.,
Legaspi,R., Maduro,Q.L., Maduro,V.B., Margulies,E.H., Masello,C.,
Maskeri,B., McDowell,J., Paguirigan,C., Pearson,R., Portnov,M.E.,
Prasad,A., Reddix-Dugue,N., Schandler,K., Schueler,M.G., Shah,K.,
Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J., Tsipouris,V.,
Vogt,J.L., Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
2 (bases 1 to 178293)
Green,E.D.
Direct Submission
Submitted (24-MAY-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
3 (bases 1 to 178293)
Green,E.D.
Direct Submission
Submitted (22-JUL-2003) NIH Intramural Sequencing Center, 8717
Government Circle, Gaithersburg, MD 20877, USA
On Jul 22, 2003 this sequence version replaced gi:32306561.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_zoo@nhgri.nih.gov
----- Project Information
```

Center project name: ecc  
Center clone name: 138L19

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig, has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 178287 bases at least Q40  
Consensus quality: 178293 bases at least Q30  
Consensus quality: 178293 bases at least Q20  
Insert size: 167000; agarose-fp  
Insert size: 178293; sum-of-contigs  
Quality coverage: 11.38x in Q20 bases; agarose-fp  
Quality coverage: 10.66x in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submittor  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 178293: contig of 178293 bp in length.

#### FEATURES

##### source

1..178293  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9823"  
/clone="RP44-138L19"  
/clone\_lib="RP44"

##### misc\_feature

1..178293  
/note="assembly\_fragment  
vector\_side:left  
clone\_end:T7  
vector\_side:right"

##### misc\_feature

1..83454  
/note="Clone overlaps with GenBank Accession Number  
AC145413 clone RP44-386P23 (center project name ecb)"

#### ORIGIN

Query Match 87.6%; Score 18.4; DB 14; Length 178293;  
Best Local Similarity 95.0%; Pred. No. 1e+02; Mismatches 0; Gaps 0;  
Matches 19; Conservative 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20

Db 156133 TGCATGTATGGGTATCTTC 156114

#### RESULT 58

AC018734/c

LOCUS AC018734 179556 bp DNA linear HTG 07-JUL-2000  
DEFINITION Homo sapiens chromosome 21 clone RP11-351D2, WORKING DRAFT  
SEQUENCE, 10 unordered pieces.

AC018734

AC018734.3 GI:8569950

KEYWORDS HTG; HTGS PHASE1; HTGS\_DRAFT.

SOURCE Homo sapiens (human)

#### ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominae; Homo.

1 (bases 1 to 179556)

Waterston,R.H.

The sequence of Homo sapiens clone

Unpublished

2 (bases 1 to 179556)

Waterston,R.H.

Direct Submission

Submitted (17-DEC-1999) Genome Sequencing Center, Washington

University School of Medicine, 444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Jun 16, 2000 this sequence version replaced gi:7023168.

#### COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml

----- Project Information -----

Center project name: H.NH0351D02

----- Summary Statistics -----

Sequencing vector: M13; 100%

Sequencing vector: plasmid; 0%

Chemistry: Dye-terminator Big Dye; 0% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 170077 bases at least Q40

Consensus quality: 173263 bases at least Q30

Consensus quality: 175196 bases at least Q20

Insert size: 180000; agarose-fp

Insert size: 178656; sum-of-contigs

Quality coverage: 5.54 in Q20 bases; agarose-fp

Quality coverage: 5.64 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 10 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2355: contig of 2355 bp in length

\* 2356 2455: gap of unknown length

\* 2456 6730: contig of 4275 bp in length

\* 6731 6830: gap of unknown length

\* 6831 17211: contig of 10381 bp in length

\* 17212 17311: gap of unknown length

\* 17312 30665: contig of 13354 bp in length

\* 30666 30765: gap of unknown length

\* 30766 45152: contig of 14387 bp in length

\* 45153 45252: gap of unknown length

\* 45253 62956: contig of 17704 bp in length

\* 62957 63056: gap of unknown length

\* 63057 79924: contig of 16868 bp in length

\* 79925 80024: gap of unknown length

\* 80025 107432: contig of 27408 bp in length

\* 107433 107532: gap of unknown length

\* 107533 138094: contig of 30562 bp in length

\* 138095 138194: gap of unknown length

\* 138195 179556: contig of 41362 bp in length.

-----

Location/Qualifiers

1..179556

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/clone="RP11-351D2"

1..2355

/note="assembly\_name:Contig6"

clone\_end:SP6

#### FEATURES

##### source

1..179556

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="21"

/clone="RP11-351D2"

1..2355

/note="assembly\_name:Contig6"

clone\_end:SP6

##### misc\_feature

1..2355

/note="assembly\_name:Contig6"

clone\_end:SP6



```

vector_side:left"
gap      2356..2455
misc_feature /estimated_length=unknown
gap      2456..6730
/notes="assembly_name:Contig7"
gap      6731..6830
/estimated_length=unknown
misc_feature 6831..17211
/notes="assembly_name:Contig8"
gap      17212..17311
/estimated_length=unknown
misc_feature 17312..30665
/notes="assembly_name:Contig9"
gap      30666..30765
/estimated_length=unknown
misc_feature 30766..45152
/notes="assembly_name:Contig10"
gap      45153..45252
/estimated_length=unknown
misc_feature 45253..62956
/notes="assembly_name:Contig11"
gap      62957..63056
/estimated_length=unknown
misc_feature 63057..79924
/notes="assembly_name:Contig12"
clone_end:T7
vector_side:right"
gap      79925..80024
/estimated_length=unknown
misc_feature 80025..107432
/notes="assembly_name:Contig13"
gap      107433..107532
/estimated_length=unknown
misc_feature 107533..138094
/notes="assembly_name:Contig14"
gap      138095..138194
/estimated_length=unknown
misc_feature 138195..179556
/notes="assembly_name:Contig15"

ORIGIN
Query Match      87.6%; Score 18.4; DB 14; Length 179556;
Best Local Similarity 95.08; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 TGCATGTATCGGTATCTTC 20
||||||| |||||
Db      158545 TGCATGTATCGATTATCTTC 158526

RESULT 59
AC118500/c
LOCUS      AC118500      211897 bp      DNA      linear      HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-144P21, WORKING DRAFT SEQUENCE, 6
unordered pieces.
ACCESSION AC118500
VERSION      AC118500.4 GI:25073691
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Rattus.
1 (bases 1 to 211897)
Muzny,D,Marle., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Ayagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Blawalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,S.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgiev,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,R.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuheva,L., Loulsegged,H., Lozado,R., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Muidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokeme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajic,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umanani,K.,
Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlarczyk,R., Woodden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 211897)
Worley,K.C.
Direct Submission
Submitted (18-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 211897)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23269807.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information

```

```

Center project name: GUIK
Center clone name: CH230-144P21
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 199093 bases at least Q40
Consensus quality: 201840 bases at least Q30
Consensus quality: 203801 bases at least Q20
Estimated insert size: 205170; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 6 contigs. The true order of the pieces
  is not known and their order in this sequence record is
  arbitrary. Gaps between the contigs are represented as
  runs of N, but the exact sizes of the gaps are unknown.
  This record will be updated with the finished sequence
  as soon as it is available and the accession number will
  be preserved.
* 1 158307: contig of 158307 bp in length
* 158308 158407: gap of unknown length
* 178808 178882: contig of 20475 bp in length
* 178883 178982: gap of unknown length
* 178983 198357: contig of 19375 bp in length
* 198358 198457: gap of unknown length
* 198458 209332: contig of 10875 bp in length
* 209333 209432: gap of unknown length
* 209433 210436: contig of 1004 bp in length
* 210437 210536: gap of unknown length
* 210537 211897: contig of 1361 bp in length.
* Location/Qualifiers
  source          1..211897
                  /organism="Rattus norvegicus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:10116"
                  /clone="CH230-144P21"
  misc_feature    1..1700
                  /note="wgs contig"
  gap             158308..158407
                  /estimated_length=unknown
  gap             178883..178982
                  /estimated_length=unknown
  misc_feature    183448..184725
                  /note="wgs contig"
  gap             198358..198457
                  /estimated_length=unknown
  gap             209333..209432
                  /estimated_length=unknown
  gap             210437..210536
                  /estimated_length=unknown
ORIGIN
Query Match      87.6%; Score 18.4; DB 14; Length 211897;
Best Local Similarity 95.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20
    |||||
Db 103579 TGCATGTATGGGTATCTTC 103560

RESULT 60
AP001751
LOCUS      Homo sapiens genomic DNA, chromosome 21q, section 95/105. PRI 21-MAY-2003
DEFINITION AP001751 AL163296 BA000005
ACCESSION  AP001751.1 GI:7768753
VERSION    .
KEYWORDS
SOURCE     Homo sapiens (human)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

```

# REFERENCE AUTHORS

Hominidae; Homo.

1 Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, K., Nagamine, K., Mitsuayama, S., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordstiek, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. and Yaspo, M.L.

The DNA sequence of human chromosome 21

Nature 405 (6784), 311-319 (2000)

10830953

2 (bases 1 to 340000)

Hattori, M., Fujiyama, A., Taylor, T.D., Watanabe, H., Yada, T., Park, H.S., Toyoda, A., Ishii, K., Totoki, Y., Choi, D.K., Soeda, E., Ohki, M., Takagi, T., Sakaki, Y., Taudien, S., Blechschmidt, K., Polley, A., Menzel, U., Delabar, J., Kumpf, K., Lehmann, R., Patterson, D., Reichwald, K., Rump, A., Schillhabel, M., Schudy, A., Zimmermann, W., Rosenthal, A., Kudoh, J., Shibuya, K., Kawasaki, K., Asakawa, S., Shintani, A., Sasaki, K., Nagamine, K., Mitsuayama, S., Antonarakis, S.E., Minoshima, S., Shimizu, N., Nordstiek, G., Hornischer, K., Barandt, P., Scharfe, M., Schoen, O., Desario, A., Reichelt, J., Kauer, G., Bloeker, H., Ramser, J., Beck, A., Klages, S., Hennig, S., Riesselmann, L., Dagand, E., Wehrmeyer, S., Borzym, K., Gardiner, K., Nizetic, D., Francis, F., Lehrach, H., Reinhardt, R. and Yaspo, M.L.

# TITLE JOURNAL PUBMED REFERENCE AUTHORS

Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic Sciences Center, Human Genome Research Group \* Institute of Molecular Biotechnology, Genome Analysis \* Keio University School of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome Analysis \* Max-Planck Institute for Molecular Genetics (addresses see below)

On May 30, 2000 this sequence version replaced gi:7717412.

The chromosome 21 mapping and sequencing consortium consisting of

\* RIKEN Genomic Sciences Center, Human Genome Research Group, \* Sagamihara 228-8555, Japan.

\* e-mail: hattori@gsc.riken.go.jp

\* URL: <http://hgp.gsc.riken.go.jp/>

and

\* Institute of Molecular Biotechnology, Genome Analysis, \* Beutenbergstrasse 11, D-07745 Jena, Germany.

\* e-mail: gscj-submit@genome.imb-jena.de

\* URL: <http://genome.imb-jena.de/>

and

\* Keio University School of Medicine, Molecular Biology, \* Tokyo 160-8582, Japan,

\* e-mail: nshimizu@dm-b-med.keio.ac.jp

\* URL: <http://www.dmb.med.keio.ac.jp/>

and

\* GBF, Dept. of Genome Analysis,

\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e-mail: info.genome@gbf.de

\* URL: <http://genome.gbf.de/>

and

\* Max-Planck Institute for Molecular Genetics,

\* Innesstrasse 73, D-14195 Berlin, Germany,

\* e-mail: info-chr21@molgen.mpg.de

\* URL: <http://chr21.rz-berlin.mpg.de/>

AL163296: Submitted (10-APR-2000).

# TITLE JOURNAL

Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic Sciences Center, Human Genome Research Group \* Institute of Molecular Biotechnology, Genome Analysis \* Keio University School of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome Analysis \* Max-Planck Institute for Molecular Genetics (addresses see below)

On May 30, 2000 this sequence version replaced gi:7717412.

The chromosome 21 mapping and sequencing consortium consisting of

\* RIKEN Genomic Sciences Center, Human Genome Research Group, \* Sagamihara 228-8555, Japan.

\* e-mail: hattori@gsc.riken.go.jp

\* URL: <http://hgp.gsc.riken.go.jp/>

and

\* Institute of Molecular Biotechnology, Genome Analysis, \* Beutenbergstrasse 11, D-07745 Jena, Germany.

\* e-mail: gscj-submit@genome.imb-jena.de

\* URL: <http://genome.imb-jena.de/>

and

\* Keio University School of Medicine, Molecular Biology, \* Tokyo 160-8582, Japan,

\* e-mail: nshimizu@dm-b-med.keio.ac.jp

\* URL: <http://www.dmb.med.keio.ac.jp/>

and

\* GBF, Dept. of Genome Analysis,

\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e-mail: info.genome@gbf.de

\* URL: <http://genome.gbf.de/>

and

\* Max-Planck Institute for Molecular Genetics,

\* Innesstrasse 73, D-14195 Berlin, Germany,

\* e-mail: info-chr21@molgen.mpg.de

\* URL: <http://chr21.rz-berlin.mpg.de/>

AL163296: Submitted (10-APR-2000).

# COMMENT

Submitted (10-APR-2000) The Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic Sciences Center, Human Genome Research Group \* Institute of Molecular Biotechnology, Genome Analysis \* Keio University School of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome Analysis \* Max-Planck Institute for Molecular Genetics (addresses see below)

On May 30, 2000 this sequence version replaced gi:7717412.

The chromosome 21 mapping and sequencing consortium consisting of

\* RIKEN Genomic Sciences Center, Human Genome Research Group, \* Sagamihara 228-8555, Japan.

\* e-mail: hattori@gsc.riken.go.jp

\* URL: <http://hgp.gsc.riken.go.jp/>

and

\* Institute of Molecular Biotechnology, Genome Analysis, \* Beutenbergstrasse 11, D-07745 Jena, Germany.

\* e-mail: gscj-submit@genome.imb-jena.de

\* URL: <http://genome.imb-jena.de/>

and

\* Keio University School of Medicine, Molecular Biology, \* Tokyo 160-8582, Japan,

\* e-mail: nshimizu@dm-b-med.keio.ac.jp

\* URL: <http://www.dmb.med.keio.ac.jp/>

and

\* GBF, Dept. of Genome Analysis,

\* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e-mail: info.genome@gbf.de

\* URL: <http://genome.gbf.de/>

and

\* Max-Planck Institute for Molecular Genetics,

\* Innesstrasse 73, D-14195 Berlin, Germany,

\* e-mail: info-chr21@molgen.mpg.de

\* URL: <http://chr21.rz-berlin.mpg.de/>

AL163296: Submitted (10-APR-2000).

# FEATURES source

1. 340000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
1. .158063

# source

1. 340000  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
1. .158063

```
source
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone_lib="Keio BAC library"
/note="Accession No. AP001046"
155173..193275
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="KUD45D10"
/clone_lib="KU21-D Cosmid library"
/note="Accession No. AP001047"
192809..209397
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="KB1216G2"
/clone_lib="Keio BAC library"
/note="Accession No. AP001048"
206177..242765
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="KUD41H11"
/clone_lib="KU21-D Cosmid library"
/note="Accession No. AP001049"
222082..321579
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="KB953G5"
/clone_lib="Keio BAC library"
/note="Accession No. AP001050"
293727..334000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="KB161B12, 3' partial"
/clone_lib="Keio BAC library"
/note="Accession No. AP001051"
1802..1827
/note="(TTCA)n"
/repeat_region
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(3874..4044)
/note="MERSA"
/rpt_family="DNA/MER1_type"
/rpt_type=DISPERSED
complement(4382..4654)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
4756..4986
/note="AluX"
/repeat_region
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
complement(5691..6047)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
```

```
repeat_region
complement(11209..11341)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(11335..11465)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
complement(11839..12177)
/note="MLT1A1"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
complement(12280..12538)
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
12539..13330
/note="(GGGGA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
13365..13608
/note="(GGGGA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
complement(15402..15494)
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
15947..16036
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
16175..16347
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
16371..16545
/note="(TCCA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
16546..16804
/note="LIMB5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
16805..17026
/repeat_region
```

Query Match 87.6%; Score 18.4; DB 8; Length 340000;  
Best Local Similarity 95.0%; Pred. No. 92;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTC 20  
|||||  
Db 90265 TGCATGTATGAGTTATCTTC 90284

Search completed: January 27, 2006, 23:13:44  
Job time : 829.191 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:46:18 ; Search time 981.18 Seconds  
(without alignments)  
1448.344 Million cell updates/sec

Title: US-10-716-005-3

Perfect score: 25

Sequence: 1 caaataaagagactattcgtcaa 25

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_in.\*

3: gb\_env.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pr.\*

9: gb\_ro.\*

10: gb\_ats.\*

11: gb\_sy.\*

12: gb\_un.\*

13: gb\_vi.\*

14: gb\_htg.\*

15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	CS109172	Sequence
2	25	100.0	1731	CQ647508	Sequence
3	25	100.0	1734	AX608629	Sequence
4	25	100.0	21781	AE014229	Streptococcus
5	25	100.0	29072	AX602188	Streptococcus
6	25	100.0	167050	1 SAG766847	Streptococcus
7	25	100.0	34980	CQ655069	Streptococcus
8	25	100.0	34980	6 AX954529	Streptococcus
9	23.4	93.6	3797	1 AB027569	Streptococcus
10	20.2	80.8	2242	1 STRPEHPO	Streptococcus
11	20.2	80.8	2259	1 STRPEHPO	Streptococcus
12	20.2	80.8	8924	1 AE014911	Streptococcus
13	20.2	80.8	110000	1 CP000029.12	Streptococcus
14	20.2	80.8	153074	14 AC163395	Mus muscu
15	20.2	80.8	194538	9 AC102008	Mus muscu
16	20.2	80.8	300029	1 AE016748	Staphylococcus
17	19.8	79.2	180692	9 AC134559	Mus muscu
18	19.8	79.2	223306	9 AC125181	Mus muscu

19	19.8	79.2	233163	14	AC113904	AC113904 Rattus no
20	19.8	79.2	235700	14	AC126298	AC126298 Rattus no
21	19.8	79.2	237969	14	AC121708	AC121708 Rattus no
22	19.8	79.2	246244	14	AC074330	AC074330 Mus muscu
23	19.8	79.2	307723	14	AC094573	AC094573 Rattus no
24	19.4	77.6	63233	8	AL359504	AL359504 Human DNA
25	19.4	77.6	164485	14	AC025340	AC025340 Homo sapi
26	19.2	76.8	11962	1	AE007858	AE007858 Clontridi
27	19.2	76.8	67233	14	AC103751	AC103751 Homo sapi
28	19.2	76.8	125319	15	AC164249	AC164249 Medicago
29	19.2	76.8	138075	14	AC149834	AC149834 Zea mays
30	19.2	76.8	138943	8	AC073928	AC073928 Homo sapi
31	19.2	76.8	156618	8	AC012362	AC012362 Homo sapi
32	19.2	76.8	176602	8	AC009306	AC009306 Homo sapi
33	19.2	76.8	182536	14	AC154981	AC154981 Bos tauru
34	19.2	76.8	182547	8	AC146141	AC146141 Pan trogl
35	19.2	76.8	183491	8	CNS01DU4	AC131163 Human chr
36	19.2	76.8	182026	9	AC157660	AC157660 Mus muscu
37	19.2	76.8	194173	8	CNS01RGW	AL160231 Human chr
38	19.2	76.8	259603	9	AC136716	AC136716 Mus muscu
39	19.2	76.8	309502	14	AC152434	AC152434 Bos tauru
40	18.8	75.2	110000	14	BX901880_2	Continuation (3 of
41	18.8	75.2	202017	5	BX323460	Continuation (3 of
42	18.8	75.2	223724	5	CR391929	CR391929 Zebrafish
43	18.8	75.2	225849	14	AC098905	AC098905 Rattus no
44	18.8	75.2	228044	5	BX682552	BX682552 Zebrafish
45	18.8	75.2	231498	14	AC120660	AC120660 Rattus no
46	18.6	74.4	575	10	BV515800	BV515800 rpf75e02.
47	18.6	74.4	694	10	BV552932	BV552932 S217P6940
48	18.6	74.4	703	10	BV529432	BV529432 GS91P6132
49	18.6	74.4	940	10	BV521567	BV521567 GS91P6618
50	18.6	74.4	1812	2	AY513464	AY513464 Stalotypa
51	18.6	74.4	2461	1	AY064171	AY064171 Streptococ
52	18.6	74.4	3991	1	AY253327	AY253327 Streptococ
53	18.6	74.4	44243	8	BS000014	BS000014 Pan trogl
54	18.6	74.4	56610	14	CR735128_4	Continuation (5 of
55	18.6	74.4	58651	14	AC017558	AC017558 Drosophil
56	18.6	74.4	69992	2	AC005421	AC005421 Drosophil
57	18.6	74.4	84055	2	AC004439	AC004439 Drosophil
58	18.6	74.4	91303	15	ATP2413	AL138655 Arabidops
59	18.6	74.4	95506	8	AC002407	AC002407 Human Chr
60	18.6	74.4	108958	8	AL451127	AL451127 Human DNA

## ALIGNMENTS

RESULT 1	CS109172	CS109172	CS109172	25 bp	DNA	linear	PAT 22-JUN-2005
LOCUS	Sequence 3 from Patent EP1541697.						
DEFINITION	CS109172						
ACCESSION	CS109172.1	GI:68148021					
VERSION							
KEYWORDS	synthetic construct						
SOURCE	other sequences; artificial sequences.						
ORGANISM	Uhl,J.R., Cockerill,F.R., Aichinger,C. and Reiser,A.						
REFERENCE	Detection of group B streptococcus						
AUTHORS	Patent: EP 1541697-A 3 15-JUN-2005;						
TITLE	Mayo Foundation for Medical Education and Research (US)						
JOURNAL	Location/Qualifiers						
FEATURES	1..25						
source	/organism="synthetic construct"						
	/mol_type="unassigned DNA"						
	/db_xref="taxon:32630"						
	/note="Oligonucleotide"						

## ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 25;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 CAAATTAAAGAGACTATTCTGCAA 25
    |||||||
Db 1 CAAATTAAAGAGACTATTCTGCAA 25

RESULT 2
CQ647508 1731 bp DNA linear PAT 02-FEB-2004
LOCUS Sequence 4465 from Patent WO0234771.
DEFINITION CQ647508
ACCESSION CQ647508
VERSION CQ647508.1 GI:41683480
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 4465 02-MAY-2002;
Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Source Location/Qualifiers
1. .1731
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCTGCAA 25
    |||||||
Db 265 CAAATTAAAGAGACTATTCTGCAA 289

RESULT 3
AX608629 1734 bp DNA linear PAT 17-FEB-2003
LOCUS Sequence 6558 from Patent WO02092818.
DEFINITION AX608629
ACCESSION AX608629
VERSION AX608629.1 GI:28404206
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS Glaser,P., Rusniok,C., Chevalier,F., Frangeul,L., Lalioui,L.,
Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and
Kunst,F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patient: WO 02092818-A 6558 21-NOV-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
Source Location/Qualifiers
1. .1734
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 25; DB 6; Length 1734;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCTGCAA 25
    |||||||
Db 265 CAAATTAAAGAGACTATTCTGCAA 289

```

RESULT 4  
AE014229  
LOCUS  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
PUBMED  
REFERENCE  
AUTHORS

TITLE  
JOURNAL

FEATURES  
source

gene  
CDS

gene  
CDS

gene

AE014229 21781 bp DNA linear BCT 23-SEP-2002  
Streptococcus agalactiae 2603V/R section 39 of 100 of the complete genome

AE014229 AE009948  
AE014229.1 GI:22533831

Streptococcus agalactiae 2603V/R  
Streptococcus agalactiae 2603V/R  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Streptococcus.

1 (bases 1 to 21781)

Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A., Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I., Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M., Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A., Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D., Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J., Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G., Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R., Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.

Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002) 12200547

2 (bases 1 to 21781)

Tettelin,H., Masignani,V., Cieslewicz,M.J., Eisen,J.A., Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I., Read,T.D., Madoff,L.C., Wolf,A.M., Beanan,M.J., Brinkac,L.M., Daugherty,S.C., DeBoy,R.T., Durkin,S., Kolonay,J.F., Umayam,L.A., Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D., Khouri,H., Mulligan,S., Carty,H.A., Cline,R.T., Gill,J., Scarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G., Mariani,M., Vegni,F., Maione,D., Rinaudo,D., Rappuoli,R., Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.

Direct Submission

Submitted (18-JUL-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. .21781  
/organism="Streptococcus agalactiae 2603V/R"  
/mol\_type="genomic DNA"  
/strain="2603V/R"  
/serotype="V"  
/db\_xref="taxon:208435"  
103. .348  
/genes="SAG0813"  
103. .348  
/notes="identified by Glimmer2; putative"

/codon\_start=1  
/transl\_table=11  
/products="hypothetical protein"  
/protein\_id="AA099700.1"  
/db\_xref="GI:22533832"  
/translation="MNNQINHCFKSIMVLGVQTMKKEFIITGRNVARVMTLVCLCMV SNLVLDLMDIIEINEIVFCVLVASIAGLEFLCKERE"

387. .674  
/genes="SAG0814"  
387. .674  
/genes="SAG0814"  
/notes="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/products="conserved hypothetical protein"  
/protein\_id="AA099701.1"  
/db\_xref="GI:22533833"  
/translation="MTYIGVPKEQKALSRRGKOLEKINOGSKTLMQIMODLVGSTCM IYLEDGSGTHPVEILKVDWDVQIRRNLRKSRKEVRIVRIHDIKELRPVKN"

719. .934

```
/gene="SAG0815"
719..934
/notes="SAG0815"
/notes="identified by match to PFAM protein family HMM
PF01381"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Cro/CI family"
/protein_id="AA099702.1"
/db_xref="GI:22533834"
/translacion="MGLIKNRKLELRARDINGNQTELAKLTGVSRTISLIERGEYTP
SVIAIKISQVFSLENVFSLVSMEE"
934..1695
/gene="SAG0816"
934..1695
/gene="SAG0816"
/notes="similar to OMNI.NTL01SPL1088; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AA099703.1"
/db_xref="GI:22533835"
/translacion="MMKNFIKQSDTTFVRIKSLKGGFIGAILSGVGAIFIIFGQDK
YLSEINIVQVFLWIVITITAFSLIYLYQIQYQKVFNFVDESQSEIYRQINLR
HSYGMPTFSISIVLSIVNTLFNFKLAIPODSVTLVPIVDLSLFLVLLGLHIYFLKVY
NTRGKIMVAPILKELKNVLQIDAELESNYKMGCFIDVMNLGSGIFFPIYFVLFFI
SFVQKVEIVITTSIHIIYIKSLKAARHFYR"
complement(1780..2343)
/gene="SAG0817"
complement(1780..2343)
/gene="SAG0817"
/notes="similar to OMNI.NTL01SPL0446; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AA099704.1"
/db_xref="GI:22533836"
/translacion="MSKNNTTCLITAIFAALAMALSMIPFASWTFPSFGAIPIL
FALRRGTGYGLWGLLHFLVLSQVFEYILAFISMLAGVFAKFKD
ALSSGKTKLALSGAILATLVRYVMHYIAGVFWYAPKMGSAITLSLVNGTA
GLTLFPVTSIIILVISYPSFPLPKK"
complement(2584..3543)
/gene="nrdrF-2"
/notes="synonym: SAG0818"
complement(2584..3543)
/gene="nrdrF-2"
/notes="identified by match to PFAM protein family HMM
PF00268"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, beta
subunit"
/protein_id="AA099705.1"
/db_xref="GI:22533837"
/translacion="MTTITYEAINNEIEDVDKSTWEKLTQFWLDRIPLSNDLDM
RKLSAQKDLGVKVGGLTLDLTQWSETGVEAIRADVRTPHEEAVALNNIOPMESVHAK
SYSISITLNTKSEIEIPEWNNNEFLQEKARIINDIYANGNALOKKVASYLETEL
FYSGFPTPLVYLGNNKLANVAEIIKLIIRDESVHGTVIGYKFLQGFNELPEDEENFR
DMWYDLYQLYENEKTKTLVDCGWTEEVMTFLRYNANKALNMLGQDPLFPDPTAND
VNPVIMGISTGTNNHFFPSQVGVNGYLLGSVEAMHDDDDYNGL"
complement(3746..5905)
/gene="nrdrE-2"
/notes="synonym: SAG0819"
complement(3746..5905)
/gene="nrdrE-2"
/notes="identified by match to PFAM protein family HMM
PF00317"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, alpha
subunit"
/protein_id="AA099706.1"
/db_xref="GI:22533838"
/translacion="MSLKNIGDVSYFRLNNEINRPVNGQIPLHKDEALTAFFKENVQ
PNSKAFDSITDKIAYLLKYDYLEEAFNLKRPDEFTELSTKLFDKKFRKFSMAAYKF
YQYALKTNDGVEYLESIEDRVFNALYFADGDEELATDALEMISORVQPATPSPLN
AGRSRGLVSLFLQVTDMMNAIGRSINSALQLSRIGGVGIGISLNLEAGAPIKGP
AAASGVVPMKLFEDSFYSNQLGQAGVYLDVDFHPDIIISFLSTKKNADKVR
VKTLSGITVPDKFYELARNNQEMLYFSYTEREYGVFYSYIDITEKYDELVPENI
TKTKINARDELETSIKLQESGYPYIINIDTANRTPNDVKIIMNSLCSEILOVQKPS
LINDAQLEMTGDISCNLSTNVLMNTSPDGKSIKTMTRALTFTVDSNIEAVPT
IKNGNAHTFGLGAMGLHSLAKNHIYEGSPESIEFTDIYFMLMNYMTLVRSNLIAR
EQQTTFVGEKSKYADGTIFYDKVSGKFPVQSDKVKSLFANHFIPAKDWNILRYAVM
KQGLYHQNKRLAVPNGSISINDCSAHPITQRIEERQEKIKIGIYYPANGLATDTI
ILRNYAFNGKRSYIYRTTDDGSEVGANQCESCVI"
complement(5983..6207)
/gene="nrdrH"
/notes="synonym: SAG0820"
complement(5983..6207)
/gene="nrdrH"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2,
NrdrH-redoxin"
/protein_id="AA099707.1"
/db_xref="GI:22533839"
/translacion="MAAITVFGKNCMQCKMTKKFLDQHGADPEEINIDEKPEKIEYV
KNLGSAPVIEAGNVFSGFQPSKLKELV"
6590..6853
/gene="ptcH"
/notes="synonym: SAG0821"
6590..6853
/gene="ptcH"
/notes="identified by match to PFAM protein family HMM
PF00381"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AA099708.1"
/db_xref="GI:22533840"
/translacion="MASKDFHIVAEIGIARPATLLVQTASKFASDITLDYKGVANL
KSIIMVMSLGVGGQADVITISAEADADDAIAAIEETWKEGLA"
6858..8591
/gene="ptcI"
/notes="synonym: SAG0822"
6858..8591
/gene="ptcI"
/notes="identified by match to PFAM protein family HMM
PF00391"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AA099709.1"
/db_xref="GI:22533841"
/translacion="MTEMLKGIASDGVAVAKAYLLVQPDLSFETVTDNASEARL
DVALQASQDLSVIREKEAEAAVFDALHMLVSDPEMINQIKETIRAKQVNAE
TGLKEVTDMEITIFEGMEDPNYQOERAADIRDVAKRVLAHLGLVCLPNPATINEESIV
IADLTPSDTAQLNKQFVKAFVTNIGRGTSHSAIMARTLEIAAVLGTNDITERVQDGO
```

Query Match 100.0%; Score 25; DB 1; Length 21781;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	CAAAATTAAGAGACTATTTCGTGCAA	25
Db	7122	CAAAATTAAGAGACTATTTCGTGCAA	7146

RESULT 5  
AX602188  
LOCUS  
DEFINITION Sequence 117 from Patent WO02092818.  
ACCESSION AX602188

VERSION AX602188.1 GI:28402057
 KEYWORDS Streptococcus agalactiae
 SOURCE Streptococcus agalactiae
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
 REFERENCE 1
 AUTHORS Glaser,P., Rusniok,C., Chevalier,F., Frangeul,L., Lalioui,L., Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Trieu-Cuot,P. and Kunat,F.
 TITLE Streptococcus agalactiae genome sequence, use for developing vaccines, diagnostic tools, and for identifying therapeutic targets
 JOURNAL Patent: WO 0292818-A 117 21-NOV-2002; INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
 FEATURES
 source
 1..29072
 /organism="Streptococcus agalactiae"
 /mol\_type="unassigned DNA"
 /db\_xref="taxon:1311"
 ORIGIN
 Query Match 100.0%; Score 25; DB 6; Length 29072;
 Best Local Similarity 100.0%; Pred. No. 3;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CAAATTAAGAGACTATTCTGTGCA 25
 Db 12160 CAAATTAAGAGACTATTCTGTGCA 12184
 RESULT 6
 SAG766847 167050 bp DNA linear BCT 16-APR-2005
 LOCUS Streptococcus agalactiae NEM316 complete genome, segment 5.
 DEFINITION AL766847 AL732656
 ACCESSION AL766847.1 GI:23095200
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM Streptococcus agalactiae NEM316
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
 REFERENCE 1
 AUTHORS Glaser,P., Rusniok,C., Buchrieser,C., Chevalier,F., Frangeul,L., Msadek,T., Zouine,M., Couve,E., Lalioui,L., Poyart,C., Trieu-Cuot,P. and Kunat,F.
 TITLE Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease
 JOURNAL Mol. Microbiol. 45 (6), 1499-1513 (2002)
 PUBMED 12354221
 REFERENCE 2
 AUTHORS Glaser,P., Rusniok,C. and Frangeul,L.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAY-2002) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68 89 96, Fax: +33 (0)1 45 68 87 86
 FEATURES
 source
 1..167050
 /organism="Streptococcus agalactiae NEM316"
 /mol\_type="genomic DNA"
 /strain="NEM316"
 /serovar="serovar III"
 /db\_xref="taxon:211110"
 join(288..293,303..1412)
 /gene="gbs0767"
 288..293
 /gene="gbs0767"
 303..1412
 /genes="gbs0767"
 /note="Similar to riboflavin specific deaminase (diaminohydroxyphosphoribosylaminopyrimidine deaminase/5-amino-6-(5-phosphoribosylamino)uracil
 gene
 3255..3739
 /gene="gbs0770"
 RBS
 3255..3260
 /gene="gbs0770"
 3269..3739
 CDS
 /note="Similar to riboflavin synthase complex beta chain"
 /codon\_start=1
 /transl\_table=11

reductase)"
 /codon\_start=1
 /transl\_table=11
 /product="Unknown"
 /protein\_id="CAD46411.1"
 /db\_xref="GI:23095201"
 /db\_xref="GOA:Q8E660"
 /db\_xref="InterPro:IPR002125"
 /db\_xref="InterPro:IPR002734"
 /db\_xref="InterPro:IPR004794"
 /db\_xref="InterPro:IPR011549"
 /db\_xref="UniProt/TREMBL:Q8E660"
 /translation="MMKEDYMALALKEAEKGMGFVAPNPLVGAVIVKDDRIISKGYH
 KFGDLHAERQAIKNAEDISGTLVVTLEPCCHVGKOPCTEALIKSGIKVVGSL
 DPNLVSGKIGALLKECLNVEVGLIRECDALNERFIFHMTYKQFPFYLYKAVTLDG
 KIATKTSKWIISNEHSQVOKRQKCSAIMVGINTVLADNPRUTCRIPKGEALVRI
 VCDSQLIPDSYLVKSARTPTWTATCSNLAAQQOQTLKMGCRILIKVPRKDKLDLK
 VLMTLIQEGIDSLIIEGSSILHPSALKAGIVNRLIVFIAPKIIIGLKAKTASGEGL
 DMLNQDFRVKDIELSRMDSDDVVEGKVEHYVVRNY"
 1381..1387
 /gene="gbs0768"
 1381..1387
 /gene="gbs0768"
 1393..2043
 /gene="gbs0768"
 1393..2043
 /gene="gbs0768"
 /note="Similar to ribiflavin synthase alpha chain"
 /codon\_start=1
 /transl\_table=11
 /product="Unknown"
 /protein\_id="CAD46412.1"
 /db\_xref="GI:23095202"
 /db\_xref="InterPro:IPR001783"
 /db\_xref="UniProt/TREMBL:Q8E659"
 /translation="MFTGIIEBMGQVSRIRNGIKSQQLSIDAPKLVLPLRKGDSDVAVN
 GVCLTVLDKSETAFIADVMPESMMRTSLAALRLHKSVALRLALRSDSRGLGHFVLGHV
 DVGKIEIKQKDDIAVRPSIDAPPSIMSYIIEKGSVALDGLSLTVVSTFTEHSEFSVI
 FHTMAQTNLKLVGDLNIEVDVLGKYAEKFLAPTNRNTHTSSVMDMSFLSENGY"
 2046..3254
 /gene="gbs0769"
 2046..2051
 /gene="gbs0769"
 2061..3254
 /gene="gbs0769"
 /note="Similar to GTP cyclohydrazase II/
 3,4-dihydroxy-2-butanone-4-phosphate synthase"
 /codon\_start=1
 /transl\_table=11
 /product="Unknown"
 /protein\_id="CAD46413.1"
 /db\_xref="GI:23095203"
 /db\_xref="GOA:Q8E658"
 /db\_xref="InterPro:IPR000422"
 /db\_xref="InterPro:IPR000926"
 /db\_xref="UniProt/TREMBL:Q8E658"
 /translation="WAFSPFKLLQDIKSGKQVVMDDENRENEGDLICAAEMVTKES
 INFMAKFGKGLICLPISNYAEKLELAQNASHTDNHETAFTISIDHLSTSGISABE
 RALTAQVANDSSKAKDFRRPGHFLPLAKKEGVLARNGHTATVDCRLAGLKECGL
 CCEIMAEQDSMMRKDELLAFAOKHDLATATIKQLQDYRQREGGVVREIEIQLPTQFG
 HETAYGVSEVAVNKEHVALVKGDISSGEDVLCRLHSECITGDFVHSLRCDCCQLANA
 LQOIEAERGVLVLYMRQGRGIGLNLKAYHLQEGELDTLEANLALGFEGERDYG
 V
 SHLLD"
 3255..3739
 /gene="gbs0770"
 3255..3260
 /gene="gbs0770"
 3269..3739
 /gene="gbs0770"
 /note="Similar to riboflavin synthase complex beta chain"
 /codon\_start=1
 /transl\_table=11



```
/product="Unknown"
/protein_id="CAD46414.1"
/db_xref="GI:23095204"
/db_xref="GOA:Q8E657"
/db_xref="InterPro:IPR002180"
/db_xref="UniProt/Swiss-Prot:Q8E657"
/translation="MTIEQLVANEMKIGIVVSRENELITSKLASGAVDGLLRHGV
SEDIDIVVPGAFIPIYMARWALYKDYDAIICLVWIKGTDHYDVVNCNEVTKGIGH
LNSQSDIPHFGVLITDNLQALIERAGTKAGNGYDCAISAEMVNLDKLRER"
/complement(3814..5315)
/gene="lysS"
/complement(3814..5304)
/gene="lysS"
/codon_start=1
/transl_table=11
/product="lysyl-tRNA synthetase"
/protein_id="CAD46415.1"
/db_xref="GI:23095205"
/db_xref="GOA:Q8E656"
/db_xref="InterPro:IPR002313"
/db_xref="InterPro:IPR004364"
/db_xref="InterPro:IPR004365"
/db_xref="InterPro:IPR006195"
/db_xref="InterPro:IPR008994"
/db_xref="UniProt/Swiss-Prot:Q8E656"
/translation="MSNOHIEELNDQIVRRKMAALTEQIDPFGKRFERTATSGOL
NEKYADKSRELDHDIETATIAGRLMTKGGKGVGAHQQDRREGQIQIYVRKDSVCEE
NYEIFKADGLDPLGVEGQVMRTDMGELSIKATHITLSKALAPLEKFGHLTDITI
YRKHDLISNRDSFDRFVTRKIISEIRFMSNGFLEVEFVLNHEAGGASARFPI
THNAQOIDMWLAIATLHLKRLIVGMERVYIGIRFNEGMDATHNPFETSIAYQ
AVADYQIDMLTGIQHVTYKVGDPGNYQGTETKINEPFKRHHVMDAVKEITGID
FWKEMTELAQAQKKNVPLEKHFTTVGHIINAFEEFVEDTLIOPTFVFGHPVVS
PLAKNDTPRFRDRELFIPTYKAYANAFELNDPIDQLSRBAQASAKELGDEATG
VDYDYVALEIGMPPGGLGIGIDRLCMLLDTITIRDLVLLFTMK"
/complement(5310..5315)
/gene="lysS"
/size .6411
/gene="gbs0772"
/size .5472
/gene="gbs0772"
/size .5479
/gene="gbs0772"
/size .6381
/notes="Similar to unknown proteins"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD46416.1"
/db_xref="GI:23095206"
/db_xref="GOA:Q8E655"
/db_xref="InterPro:IPR005833"
/db_xref="InterPro:IPR005834"
/db_xref="InterPro:IPR006439"
/db_xref="UniProt/TrEMBL:Q8E655"
/translations="MITSIVFDVDDTIYDQAPYIAVEKCFDFDMSAINQAYIRFR
HYSIDPLPRVAGETWTEYFRFWRKETLLEFGVEIDEATGIYFQBIYSHLENTIM
LDEMWTLLFSKNVPMGIITNGPTEHQLKKVKGLGLDYDVPKRVIVSQATGFOKPK
EKGIFFNLAAQFQDNSTLYVGDSYDNDINGAFNGGWHSMFNFNRCRSRKPKGIKPYI
VDKIDNFEQLFGAVKVLFDLPDNKRFIDINDKSNPVLNGLNGLMAAERGLLESNMS
VDKVVILLRLTAQEKVLRMKYAR"
/size .6411
/gene="gbs0772"
/complement(6417..7069)
/gene="gbs0773"
/complement(6417..7058)
/gene="gbs0773"
/notes="Similar to putative phosphoglycerate mutase"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD46417.1"
/db_xref="GI:23095207"
/db_xref="GOA:Q8E654"
/db_xref="InterPro:IPR001345"
```

```
/db_xref="UniProt/TrEMBL:Q8E654"
/translation="MKFYLVRRHGKTMNLEGRFQGGANGSDPLLEBAIELEBELGQVLS
SIHFDAVYSSDLGRADTVNILDANSQCPKEIHYTPQLREWALGTLEGCKIATMOAIY
PRQMTAFYQNPLOPKHDMFGAESLYQTHRVESFLRSLASKNYDKVLIYVGHGANLTAS
IRSLGQYQGSLSHYKDKLDNASLTIIETHDFKDFNCLTWNDKSYLRQEVQMTM"
/complement(7064..7069)
/gene="gbs0773"
/complement(7080..7553)
/gene="gbs0774"
/complement(7080..7553)
/gene="gbs0774"
/notes="Similar to transcriptional regulator"
/codon_start=1
/transl_table=11

Query Match 100.0%; Score 25; DB 1; Length 167050;
Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Gaps 0;
Matches 25; Conservative 0; Indels 0;

QY 1 CAAATTAAAGAGACTATTCGTGCAA 25
|||||
Db 79926 CAAATTAAAGAGACTATTCGTGCAA 79950

RESULT 7
CQ655069 349980 bp DNA linear PAT 02-FEB-2004
Sequence 12026 from Patent WO0234771.
DEFINITION CQ655069
ACCESSION CQ655069
VERSION CQ655069.1 GI:41687946
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE 1
AUTHORS Telford,J., Masignani,V., Margarit y Ros,I., Grandi,G., Fraser,C.
and Tettelin,H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 12026 02-MAY-2002;
Chiron S.p.A. (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
Location/Qualifiers
source
1..349980
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"
/notes="seq 10967 to big: splitted in
seq 10.967, from 0.000.001 to 0.349.980
seq 12.025, from 0.300.001 to 0.649.980
seq 12.026, from 0.600.001 to 0.949.980
seq 12.027, from 0.900.001 to 1.249.980
seq 12.028, from 1.200.001 to 1.549.980
seq 12.029, from 1.500.001 to 1.849.980
seq 12.030, from 1.800.001 to 2.149.980
seq 12.031, from 2.100.001 to 2.160.266"

ORIGIN

Query Match 100.0%; Score 25; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 1.8; Mismatches 0; Gaps 0;
Matches 25; Conservative 0; Indels 0;

QY 1 CAAATTAAAGAGACTATTCGTGCAA 25
|||||
Db 191703 CAAATTAAAGAGACTATTCGTGCAA 191727

RESULT 8
AX954529 349980 bp DNA linear PAT 08-JAN-2004
LOCUS AX954529
DEFINITION Sequence 1375 from Patent WO03093306.
ACCESSION AX954529
VERSION AX954529.1 GI:40783902
KEYWORDS
```

```

SOURCE      Streptococcus pyogenes
ORGANISM     Streptococcus pyogenes
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE    1
AUTHORS      Telford,J., Maignani,V., margarit y Ros,I., Grandi,G., Fraser,C.
             and Tettelin,H.
TITLE        Nucleic acids and proteins from streptococcus groups a b
JOURNAL      Patent: WO 03093306-A 1375 13-NOV-2003;
             Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES     Location/Qualifiers
             1..349980
               /organism="Streptococcus pyogenes"
               /mol_type="unassigned DNA"
               /db_xref="taxon:1314"
               /notes="seq 1373 to long, original length: 2.160.266
               replaced by
               seq 1373: from 0.000.001 to 0.349.980
               seq 1374: from 0.300.001 to 0.649.980
               seq 1375: from 0.600.001 to 0.949.980
               seq 1376: from 0.900.001 to 1.249.980
               seq 1377: from 1.200.001 to 1.549.980
               seq 1378: from 1.500.001 to 1.849.980
               seq 1379: from 1.800.001 to 2.149.980
               seq 1380: from 2.100.001 to 2.160.266"
ORIGIN
Query Match      100.0%; Score 25; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 CAAATTAAAGAGACTATTCTGTCGAA 25
        |||||
Db      220490 CAAATTAAAGAGACTATTCTGTCGAA 220514

RESULT 9
AB027569 LOCUS      3797 bp DNA linear BCT 23-JUL-2003
DEFINITION Streptococcus bovis ptsH, ptsI genes for histidine containing
             protein, phosphoenolpyruvate-protein phosphotransferase, complete
             cds.
ACCESSION AB027569
VERSION    3 GI:67232275
KEYWORDS   phosphoenolpyruvate-protein phosphotransferase; histidine
             containing protein.
SOURCE     Streptococcus bovis
ORGANISM   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE  1
AUTHORS    Asanuma,N. and Hino,T.
TITLE      Molecular characterization of Hpr and related enzymes, and
             regulation of Hpr phosphorylation in the ruminal bacterium
             Streptococcus bovis
JOURNAL    Arch. Microbiol. 179 (3), 205-213 (2003)
PUBMED     12610726
REFERENCE  2 (bases 1 to 3797)
AUTHORS    Asanuma,N. and Hino,T.
TITLE      Direct Submission
JOURNAL    Submitted (24-MAY-1999) Narito Asanuma, Meiji University,
             Department of Agriculture; Higashimita, Tama-ku, Kawasaki, Kanagawa
             214-8571, Japan (E-mail:asanuma@isc.meiji.ac.jp,
             Tel:+81-44-934-7825 (ex.7825), Fax:+81-44-934-7825)
COMMENT    On Jan 20, 2000 this sequence version replaced gi:5706360.
FEATURES   Location/Qualifiers
             1..3797
               /organism="Streptococcus bovis"
               /mol_type="genomic DNA"
               /strain="JBI"
               /db_xref="taxon:1315"
               1779..2042
               /gene="ptsH"

SOURCE      Streptococcus pyogenes
ORGANISM     Streptococcus pyogenes
             Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE    1
AUTHORS      Telford,J., Maignani,V., margarit y Ros,I., Grandi,G., Fraser,C.
             and Tettelin,H.
TITLE        Nucleic acids and proteins from streptococcus groups a b
JOURNAL      Patent: WO 03093306-A 1375 13-NOV-2003;
             Chiron SRL (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES     Location/Qualifiers
             1..349980
               /organism="Streptococcus pyogenes"
               /mol_type="unassigned DNA"
               /db_xref="taxon:1314"
               /notes="seq 1373 to long, original length: 2.160.266
               replaced by
               seq 1373: from 0.000.001 to 0.349.980
               seq 1374: from 0.300.001 to 0.649.980
               seq 1375: from 0.600.001 to 0.949.980
               seq 1376: from 0.900.001 to 1.249.980
               seq 1377: from 1.200.001 to 1.549.980
               seq 1378: from 1.500.001 to 1.849.980
               seq 1379: from 1.800.001 to 2.149.980
               seq 1380: from 2.100.001 to 2.160.266"
ORIGIN
Query Match      93.6%; Score 23.4; DB 1; Length 3797;
Best Local Similarity 96.0%; Pred. No. 21;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1 CAAATTAAAGAGACTATTCTGTCGAA 25
        |||||
Db      2311 CAAATTAAAGAACTATTCTGTCGAA 2335

RESULT 10
STRPHOSPHO LOCUS      2242 bp DNA linear BCT 19-OCT-2000
DEFINITION Streptococcus mutans phosphoenolpyruvate:sugar phosphotransferase
             system Hpr (ptsH) gene, complete cds; phosphoenolpyruvate:sugar
             phosphotransferase system enzyme I (ptsI) gene, complete cds.
ACCESSION L15191.1 GI:310626
VERSION    1
KEYWORDS   phosphoenolpyruvate:sugar phosphotransferase system Hpr;
             Streptococcus mutans
SOURCE     Streptococcus mutans
ORGANISM   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

REFERENCE  1 (bases 1 to 2242)
AUTHORS    Boyd,D.A., Cvitkovitch,D.G. and Hamilton,I.R.
TITLE      Sequence and expression of the genes for Hpr (ptsH) and enzyme I
             (ptsI) of the phosphoenolpyruvate-dependent phosphotransferase
             transport system from Streptococcus mutans
JOURNAL    Infect. Immun. 62 (4), 1156-1165 (1994)
PUBMED     8132321
REFERENCE  2 (bases 1 to 2242)
AUTHORS    Boyd,D.A., Cvitkovitch,D.G. and Hamilton,I.R.
TITLE      Sequence, expression, and function of the gene for the
             nonphosphorylating, NADP-dependent glyceraldehyde-3-phosphate
             dehydrogenase of Streptococcus mutans
JOURNAL    J. Bacteriol. 177 (10), 2622-2627 (1995)
PUBMED     7751269
FEATURES   Location/Qualifiers
             1..2242
               /organism="Streptococcus mutans"
               /mol_type="genomic DNA"
               /strain="NG5"

```



JOURNAL  
PUBMED  
REFERENCE  
AUTHORS  
Ajdic,D., McShan,W.M., McLaughlin,R.E., Savic,G., Chang,J.,  
Carson,M.B., Prineaux,C., Tian,R., Kenton,S., Jia,H., Lin,S.,  
Qian,Y., Li,S., Zhu,H., Najjar,F., Lai,H., White,J., Roe,B.A. and  
Ferretti,J.J.  
Direct Submission  
Submitted (09-JUL-2002) Department of Microbiology and Immunology,  
University of Oklahoma Health Sciences Center, 940 SL Young Blvd.,  
Oklahoma City, OK 73104, USA  
FEATURES  
Location/Qualifiers  
1. .8924  
/organism="Streptococcus mutans UA159"  
/mol\_type="genomic DNA"  
/strain="UA159"  
/db\_xref="taxon:210007"  
365. .1162  
/locus\_tag="SMU.673"  
365. .1162  
/locus\_tag="SMU.673"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAN58407.1"  
/db\_xref="GI:24377052"  
/translation="MWISINLVFLFFYSFTGWLWETVYCSLKDKQFYVRGFLAGPY  
CPVYFAVTVLFTKQFQENLPILFFGLIIATFEYMAACFLKVFHKLWDYSQE  
SNGIKRIPALSLFGWGVLFKFIQPSVNMQLAKNANWALGALFMTADFMTV  
SDIASFORAALSLEKMSAGKDIOAYLREIWNLEEDTDFYLSLRQFLHQLIEKMR  
RGIQPFPHQRMLRNYHLKLKDAPFAVAAVHKLASVKNKDKKR"  
1328. .1591  
/gene="ptsh"  
/locus\_tag="SMU.674"  
1328. .1591  
/gene="ptsh"  
/locus\_tag="SMU.674"  
/notes="Best Blastp Hit: sp|P45596|PTHP\_STRMU  
PHOSPHOCARRIER PROTEIN HPR (HISTIDINE-CONTAINING PROTEIN)  
>gi|482380|pir||A44562 phosphotransferase system  
phosphohistidine-containing protein - Streptococcus mutans  
>gi|310627|gb|AA91092.1| (L15191)  
phosphoenolpyruvate:sugar phosphotransferase system Hpr  
[Streptococcus mutans]"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoenolpyruvate:sugar phosphotransferase  
system Hpr"  
/protein\_id="AAN58408.1"  
/db\_xref="GI:24377053"  
/translation="WASKDFHIVAEETGHIHARPATLLVQTASKFASDITLDYKGAVNL  
KSIQVMSLGVGGADVTITABGADADDAIRAINETMTKEGLA"  
1596. .3329  
/locus\_tag="SMU.675"  
1596. .3329  
/locus\_tag="SMU.675"  
/notes="Best Blastp Hit: sp|P45595|PTL\_STRMU  
PHOSPHOENOLPYRUVATE-PROTEIN PHOSPHOTRANSFERASE  
(PHOSPHOTRANSFERASE SYSTEM, ENZYME I)  
>gi|310628|gb|AA91093.1| (L15191)  
phosphoenolpyruvate:sugar phosphotransferase system enzyme  
I [Streptococcus mutans]"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoenolpyruvate:sugar phosphotransferase  
system enzyme I, pts system EI component"  
/protein\_id="AAN58409.1"  
/db\_xref="GI:24377054"  
/translation="WTEMLKGIATAADSGVAVAKAYLLVQPDLTFTVSVTDTQAEARL  
DAALKEASQNSLIRKQAVDTLGEAAAVFDALHVLADPEMIGQIKETIRTKYNVAE  
SALKVEITDMFVLTFENMEDNPNYQEAADIRDAKRVLAHLGLVLPNPAITSEESIV  
IAHDLTPSTAOLDANYKFAVTNIGGRTSHSAIIVARTLEIAAVLGTNDITBRVKNGD  
IVAVGITTQVIINPTEDQIAEFKAAGETIYAKQKAEWALLKDAETVTADGKHFEAAN

IGTPKDVGVNNGAEAVGLYRTEFLYMDSDPFTTEDEQYEAYKAVLEGNGKPYVVR  
TMDIGDKELPYFDLPKEMNPFGLFRALRISISSETGNOMPTQLRALRASVHGQLRIV  
NFMVALLNEFRKAKGILBEEKANLKAEGVAVSDDIQVIMIEIPAAAMLADQPAKEV  
DFFSIGTNDLLQYTMADRMNEQYSYLYQPNPSILRLVDHVHVAHAEGKAGNCGE  
MAGDOTAVPLLVGLGDFEFSMSATSVLRTSLMKKLDPTDKMQLAQRALTTCATMEEV  
LELEKEVIDP"  
3559. .4986  
/gene="gapN"  
/locus\_tag="SMU.676"  
3559. .4986  
/gene="gapN"  
/locus\_tag="SMU.676"  
/EC\_number="1.2.1.9"  
/note="Best Blastp Hit: sp|Q59931|GAPN\_STRMU  
NADP-DEPENDENT GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE  
(NON-PHOSPHORYLATING GLYCERALDEHYDE 3-PHOSPHATE  
DEHYDROGENASE) (GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE  
[NADP+]) (TRIOSEPHOSPHATE DEHYDROGENASE)  
>gi|2117515|pir||A57151 glyceraldehyde-3-phosphate  
dehydrogenase (NADP+) (EC 1.2.1.9) - Streptococcus mutans  
>gi|5822059|pdb|1EUH|A Chain A, Apo Form Of A NADp  
Dependent Aldehyde dehydrogenase From Streptococcus Mutans  
>gi|5822060|pdb|1EUH|B Chain B, Apo Form Of A NADp  
Dependent Aldehyde dehydrogenase From Streptococcus Mutans  
>gi|5822061|pdb|1EUH|C Chain C, Apo Form Of A NADp  
Dependent Aldehyde dehydrogenase From Streptococcus Mutans  
>gi|5822062|pdb|1EUH|D Chain D, Apo Form Of A NADp  
Dependent Aldehyde dehydrogenase From Streptococcus Mutans  
>gi|5822470|pdb|2EUH|A Chain A, Holo Form Of A NADp  
Dependent Aldehyde dehydrogenase Complex With NADp+  
>gi|5822471|pdb|2EUH|B Chain B, Holo Form Of A NADp  
Dependent Aldehyde dehydrogenase Complex With NADp+  
>gi|5822472|pdb|2EUH|C Ch"  
/codon\_start=1  
/transl\_table=11  
/product="NADP-dependent glyceraldehyde-3-phosphate  
dehydrogenase"  
/protein\_id="AAN58410.1"  
/db\_xref="GI:24377055"  
/translation="WTKQKNYNGEWKLSNEIKIYEPASGAELGSPVPMSTEVDY  
VYAKKAQAPARSLSYIERAAAYLHKVADILMDKDKETGAVLSKEVAGYSVSEVV  
RTAETINYAABEGLMEGEVLEGGFSAADSKKTAVVRPEVGLVLAISPYNPNLA  
EQKAPALIGNAVTAFKPTQGSIGLLAFAEAGLPAGVNFTITGRSGSIGYIV  
EHQAVFINFTSGTIGIGRICKMAGMRPIMLELGGKSAIVLEADADLTAKNITAGA  
FGYSGQRCTAVKRVLMESVADELVEKIREKVALTICNPEDDADITPLDITKSADYV  
EGLINDKNGAALATEIKRGNLICPILFDKVTYDMRLANKEEPGPVLPIRVTSE  
BAIISNKSSEYGLQASIFTNDFPRAFGIAEQLEVGTVHNNKTQKGTDNFPLGAKKS  
GAGIQGVKYSIEAMTTKSVVFDIK"  
5657. .6049  
/locus\_tag="SMU.677"  
5657. .6049  
/locus\_tag="SMU.677"  
/note="Best Blastp Hit: pir||B69970 transcription  
regulator MerR family homolog yraB - Bacillus subtilis  
>gi|2108269|emb|CAA63468.1| (X92868) mercuric resistance  
operon regulatory protein [Bacillus subtilis]  
>gi|2635146|emb|CAB14642.1| (Z99117) similar to  
transcriptional regulator (MerR family) [Bacillus  
subtilis]"  
/codon\_start=1  
/transl\_table=11  
/product="putative transcriptional regulator (MerR  
family)"  
/protein\_id="AAN58411.1"  
/db\_xref="GI:24377056"  
/translation="WNIKKVSEITDISADTIYRIGLLPRITRNSGVDRDFTREI  
GILFVRCFRKAGMSVEALIEVISLBERGETERERLLLTQORDMDRDIYELNQAR  
ERLANYKIENYENIIQKHQKLFTESEEN"  
6049. .6891  
/locus\_tag="SMU.678"  
6049. .6891  
/locus\_tag="SMU.678"  
/note="Best Blastp Hit: pir||A72308 oxidoreductase,

aldo/keto reductase family - Thermotoga maritima (strain MSB8) >gi|4981536|gb|AAD36074.1|AE001762.1 (AE001762) oxidoreductase, aldo/keto reductase family [Thermotoga maritima]  
/codon\_start=1  
/transl\_table=11  
/product="putative oxidoreductase, aldo/keto reductase family"  
/protein\_id="AAN58412.1"  
/db\_xref="GI:24377057"  
/translation="METVVLNSGVKMPILGFGVQVDPDLBECERVVTEAIKVGYRLLD  
TAQVGNESAVGNARSKGLAEFEFTTKWISNAGYEKAKASLDLSRUKQDTYVD  
LVLIHQFDGYTGTSALAEYKSKAIGVSNFYFDRLVLDLASFNEIVPMVNQVET  
HYFNOIQIEAKYSEEYGVQLESWGPFAEKGNNFTNPTLTATIGEKYKSAAGVGLRYL  
IORGVVIPKTVHQRWQNFVDVFGNLSDDDDMADILKLDQGESLFINHYDPTVKFL  
ANYK"  
6901. .7767  
/locus\_tag="SMU.679"  
6901. .7767  
/locus\_tag="SMU.679"  
/notes="Best Blastp Hit: pir||A72308 oxidoreductase, aldo/keto reductase family - Thermotoga maritima (strain MSB8) >gi|4981536|gb|AAD36074.1|AE001762.1 (AE001762) oxidoreductase, aldo/keto reductase family [Thermotoga maritima]"  
/codon\_start=1  
/transl\_table=11  
/product="putative oxidoreductase, aldo/keto reductase family"  
/protein\_id="AAN58413.1"  
/db\_xref="GI:24377058"  
/translation="MEYITLNTGVKMPMEGFGVQVTDPEICKESVLNAIKTGYYRLF  
D  
TAAYGNEXALGEAHEAIEQGLVTRNELFITSKLWQDMYNEAAKKGIERSLEKGL  
DYLDLYLHQARDYEGYALAEAYQAGKRAIGVSNFYFNILTNFCTVSVIPAVN  
QIELHFFQTEALEATYRYGVVPEAPLGGGYDAFNNAILQVESAHQKTVQGM  
LWNVQRGVVVIPIKSTHKRIENFAINDFSLTDDMTKIASLDMGYGDTRTKHFDPN  
FVRSVLNVKIID"

gene 6901. .7767  
CDS 6901. .7767  
/locus\_tag="SMU.679"  
/notes="Best Blastp Hit: pir||A72308 oxidoreductase, aldo/keto reductase family - Thermotoga maritima (strain MSB8) >gi|4981536|gb|AAD36074.1|AE001762.1 (AE001762) oxidoreductase, aldo/keto reductase family [Thermotoga maritima]"  
/codon\_start=1  
/transl\_table=11  
/product="putative oxidoreductase, aldo/keto reductase family"  
/protein\_id="AAN58413.1"  
/db\_xref="GI:24377058"  
/translation="MEYITLNTGVKMPMEGFGVQVTDPEICKESVLNAIKTGYYRLF  
D  
TAAYGNEXALGEAHEAIEQGLVTRNELFITSKLWQDMYNEAAKKGIERSLEKGL  
DYLDLYLHQARDYEGYALAEAYQAGKRAIGVSNFYFNILTNFCTVSVIPAVN  
QIELHFFQTEALEATYRYGVVPEAPLGGGYDAFNNAILQVESAHQKTVQGM  
LWNVQRGVVVIPIKSTHKRIENFAINDFSLTDDMTKIASLDMGYGDTRTKHFDPN  
FVRSVLNVKIID"

Query Match 80.8%; Score 20.2; DB 1; Length 8924;  
Best Local Similarity 88.0%; Pred. No. 3.7e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCAA 25  
Db 1860 CAGATTAAGAGACAATTCTGACAA 1884

RESULT 13  
CP000029\_12/c  
WPCOMENT  
Sequence split into 27 fragments LOCUS CP000029 Accession CP000029

Fragment Name	Begin	End
CP000029_00	1	110000
CP000029_01	100001	210000
CP000029_02	200001	310000
CP000029_03	300001	410000
CP000029_04	400001	510000
CP000029_05	500001	610000
CP000029_06	600001	710000
CP000029_07	700001	810000
CP000029_08	800001	910000
CP000029_09	900001	1010000
CP000029_10	1000001	1110000
CP000029_11	1100001	1210000
CP000029_12	1200001	1310000
CP000029_13	1300001	1410000
CP000029_14	1400001	1510000
CP000029_15	1500001	1610000
CP000029_16	1600001	1710000
CP000029_17	1700001	1810000
CP000029_18	1800001	1910000
CP000029_19	1900001	2010000
CP000029_20	2000001	2110000

CP000029\_21 2100001 2210000  
CP000029\_22 2200001 2310000  
CP000029\_23 2300001 2410000  
CP000029\_24 2400001 2510000  
CP000029\_25 2500001 2610000  
CP000029\_26 2600001 2616530  
Continuation (13 of 27) of CP000029 from base 1200001 (CP000029 Staphylococcus epidermi.

Query Match 80.8%; Score 20.2; DB 1; Length 110000;  
Best Local Similarity 88.0%; Pred. No. 2.3e+02;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCAA 25  
Db 79426 CAAATTAAGAGAAAATTCTGCAA 79402

RESULT 14  
AC163395/c  
LOCUS AC163395 153074 bp DNA linear HTG 12-JUL-2005  
DEFINITION Mus musculus chromosome 7 clone RP24-251J14 map 7, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 4 unordered pieces.  
AC163395  
VERSION HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 153074)  
Birten,B., Nuebaum,C. and Lander,E.  
Mus musculus chromosome 7, clone RP24-251J14  
Unpublished  
2 (bases 1 to 153074)  
Birten,B., Nuebaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Miengua,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (10-JUN-2005) Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 153074)  
Birten,B., Nuebaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,M., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Liu,X., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Miengua,V., Murphy,T., Naylor,J., Nguyen,C., Nguyen,T., Nicol,R., Norbu,C.,



Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupback, R., Seaman, S., Sever, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Teffaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-MAR-2004) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 30, 2004 this sequence version replaced gi:42544081.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@broad.mit.edu](mailto:sequence.submissions@broad.mit.edu)

----- Project Information

Center project name: L17869

Center clone name: 502\_L\_8

-----

## FEATURES

## source

Location/Qualifiers

1..194538

/organism="Mus musculus"

/mol\_type="Genomic DNA"

/db\_xref="taxon:10090"

/chromosomes="7"

/map="7"

/clone="RP24-502L8"

/clone\_lib="RPCI-24 Male Mouse BAC"

## misc\_feature

1..14504

/note="wgs end extension"

/clone="SP6"

/complement(363..477)

/rpt\_family="ORR1D"

638..678

/rpt\_family="(CA)n"

## repeat\_region

817..1267

/rpt\_family="IAPLTR2\_MM"

2069..2121

/rpt\_family="(CA)n"

2686..2827

/rpt\_family="GA-rich"

## repeat\_region

/complement(4211..4567)

/rpt\_family="MTD"

/complement(4736..4856)

## unsure

/note="single clone coverage"

/complement(4753..4757)

## unsure

/note="<30 qual single clone coverage"

/complement(4781..4884)

## repeat\_region

4888..4910

/rpt\_family="LIME"

## repeat\_region

5004..5073

/rpt\_family="(CA)n"

## unsure

/note="single clone coverage"

5264..5354

## repeat\_region

/rpt\_family="B2\_Mm2"

5415..5617

## repeat\_region

/rpt\_family="B3"

5619..5673

## repeat\_region

/rpt\_family="(CAT)n"

6467..6495

## repeat\_region

/rpt\_family="AT rich"

/complement(6509..6586)

## repeat\_region

/rpt\_family="IDS"

/complement(6595..6698)

## repeat\_region

/rpt\_family="PB1D7"

7205..7235

## repeat\_region

/rpt\_family="AT rich"

7611..7730

## repeat\_region

/rpt\_family="B3A"

repeat\_region 7769..7799

/rpt\_family="(CA)n"

repeat\_region 8737..8781

/rpt\_family="(GGAA)n"

repeat\_region 8782..8890

/rpt\_family="B1P"

repeat\_region 8897..9090

/rpt\_family="B2\_Mm2"

repeat\_region 10500..10764

/rpt\_family="Lx8"

repeat\_region 11216..11347

/rpt\_family="B1\_MM"

repeat\_region 11370..11403

/rpt\_family="(CAAAAA)n"

repeat\_region complement(11558..11715)

/rpt\_family="MTE"

repeat\_region complement(12465..12587)

/rpt\_family="MIR"

repeat\_region 12674..12721

/rpt\_family="(TG)n"

repeat\_region complement(13787..13900)

/rpt\_family="Lx2"

misc\_feature 14905..14908

/note="clone boundary"

/clone\_end:SP6

site:MboI"

repeat\_region 15839..15898

/rpt\_family="(TCTA)n"

repeat\_region 16060..16224

/rpt\_family="MTE"

repeat\_region complement(16898..17123)

/rpt\_family="ORR1D"

repeat\_region complement(17144..17338)

/rpt\_family="ORR1D"

repeat\_region complement(17456..17604)

/rpt\_family="RSINE1"

repeat\_region complement(17732..17768)

/rpt\_family="B2\_Mm2"

repeat\_region 17769..17797

/rpt\_family="(TTTA)n"

repeat\_region complement(17798..17951)

/rpt\_family="B2\_Mm2"

repeat\_region complement(17968..18117)

/rpt\_family="B1\_MM"

repeat\_region complement(19330..19518)

/rpt\_family="LIMC2"

repeat\_region complement(19712..19891)

/rpt\_family="LIMC2"

Query Match 80.8%; Score 20.2; DB 9; Length 194538;

Best Local Similarity 88.0%; Pred. No. 2e+02; Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTCAA 25

|||||

Db 118542 CAAATTAAGAGACTATTCGTCAA 118518

## RESULT 16

## AE016748/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

AE016748 300029 bp DNA linear BCT 22-APR-2003

Staphylococcus epidermidis ATCC 12228, section 5 of 9 of the complete genome.

AE016748 AE015929

AE016748.1 GI:27315631

Staphylococcus epidermidis ATCC 12228

Staphylococcus epidermidis ATCC 12228

Bacteria; Firmicutes; Bacillales; Staphylococcus.

1 (bases 1 to 300029)

Zhang, Y., Ren, S., Li, H., Fu, G., Lu, L., Lu, G., Jia, J., Tu, Y., Qin, Z., Chen, Z. and Wen, Y.

Direct Submission



JOURNAL Submitted (05-NOV-2002) Chinese National Human Genome Center at Shanghai, 250 Bi Bo Road, Shanghai 201203, China

FEATURES

source

1..300029

/organism="Staphylococcus epidermidis ATCC 12228"

/mol\_type="genomic DNA"

/strain="ATCC 12228"

/db\_xref="ATCC:12228"

/db\_xref="taxon:176280"

complement(137..1123)

/locus\_tag="SE1168"

complement(137..1123)

/locus\_tag="SE1168"

/codon\_start=1

/transl\_table=11

/product="thioredoxin reductase-like protein"

/protein\_id="AAO04767.1"

/db\_xref="GI:27315632"

/translation="MQTIESIIIGGPGCLSAAIISQKKGIETLVIEKGNVVESTIYNY

PTHQFPSSDKLSGDIFFVEDSKPRNQALVYREVVKHQNLNHPFEVLTVKK

INNKFAITTKGVECKYLTVAITYGQHTLEAGSELPVHFYFKAHPFNQNVV

IIGKNSADALEKAGANVTLYRGEYPAIKPWLNFESLVNHEKITMEFNA

TWTKTDSVYKDGQLEIENDYVFAMIGYHPDYDFLKTIGIDHTNEXGTAPVYN

RETPTNVENCYIAGVIAAGNDANTIFIENGYHGVITQSILTKKQPLET"

complement(1256..2674)

/locus\_tag="SE1169"

complement(1256..2674)

/locus\_tag="SE1169"

/codon\_start=1

/transl\_table=11

/product="elastin binding protein"

/protein\_id="AAO04768.1"

/db\_xref="GI:27315633"

/translation="MIEIFKMGMIIVSNNNFKDFEKNRQSNINDEQOTELKEDDKT

NEKKEASQNSLNSNQOFPFPRNAQRKRRETATNQSKQDDKHQKNSDAKTTEG

SLDDYDAQLQQHDKSQONKTEKQSDNRMKGDAAI VNGTSESPEHKSSTON

RPGRQAQQKRSESTQSPSTNKKATGAGIAGAVAGAETSKRHHNKDKQD

SKSHENDERSVKNDQKQSKKGAAGVAGAAAGVGAHNNQNKHNEKNS

NONQNYNQSEGKGGKMKILLPIAAILLGAIAIFGGMALNNHNSKSDQKIAN

QSKDKSDQAQSDNDKQSDNKKKSDSDNADDDSDNSSNPATSTNNNDNY

ANNNGNYTNQONNANQSNQATQQQSHTVYQENLYRIAIIQYTGEGTQANVVK

IKRANGLSNNIHTNGQTVIQQ"

complement(2795..4174)

/locus\_tag="SE1170"

complement(2795..4174)

/locus\_tag="SE1170"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAO04770.1"

/db\_xref="GI:27315634"

/translation="MNRLNSVYIKMLTILIVATVLVLLSVSLITFIQYFTQETEN

AIKEDAKRISSLVKSHNKTIAIQNSQTLIDGPGGLIIMKMSKDPSSYKDTKKQ

MFNEIKKSTKFKYKEGETQNTITIKNGNSQSYLLGLYPMKAQGAQSHYSGVFI

YKDLKSTEDTNNAITIIITAIITIASTIAPFLSNRIKPLRQLTKQAQKVSQEGD

YSQISTVATKDEIGDLSAFNNMVEIQEHKAISSKNIRDTLLNSMVEGLGNNO

REILSNKMADDIMRHDDPFKSEIEQIEQIEATFESQOQNEYLELEINTRYVFISSYID

RIQNGRSQGIYVMVIRDMTNHNLQMKKDFIANVSHELURTPIQLQHTESLVDGIVT

EPDETIRSLAIVLDESRLNRLNVELNVARMDAEGLSVEKELQPIQLHLLQMSKSYR

MQSELGLTMTFDSNDEQLWNVDMDQVLTNLIDNATRYTQAGDSIKISIDSDSD

NIILITDTGTGIAPEHLKQVDFYKVDAAARKKGQCTGLGLFCIKMIIIEHGGRID

VESLKGKTSFIILPKSKQIS"

complement(8766..9491)

/locus\_tag="SE1176"

complement(8766..9491)

/locus\_tag="SE1176"

/codon\_start=1

/transl\_table=11

/product="staphylococcal respiratory response protein

SrrA"

/protein\_id="AAO04775.1"

/db\_xref="GI:27315640"

/translation="WTNEILIVDDDEDIRRLKLYLRESFEIHEARDKEAYELALE

NNYACILDLMLPEMDGIEVASKREHKDTIIMLTAKGENTNRVEGESGADDYIVK

PFSPREVVLYRKALLRRQTANTQSEPHARDIIEFSLVTDNDARRVLDDQQVNLIT

PKYELVLRVLAKTNNKTRFQEBQQLKVVHYEFYGLDRTVDTHVKKLRKLNKRVSSDA

QMIQTVMGVGYKFEVPSDEPTK"

complement(9625..10362)

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

LNTYLYQKLSGLNLKEIAQLQONKLVNTIEDHILEMYIKGYLIDYTLFINKKIDLE

FINYVQKHGRGERLAKFYKEHFDWTYFQIKLVIGIERGDLIAER"

5232..5480

/locus\_tag="SE1172"

5232..5480

/locus\_tag="SE1172"

/codon\_start=1

/transl\_table=11

/product="ferredoxin"

/protein\_id="AAO04771.1"

/db\_xref="GI:27315636"

/translation="MAKYTIVDMDTCIACGAGAAAPDIYDDDEGIAYVILDDNQGT

AEVPELYEDMEDALEKPTDSIKIIEDEPFQGDALKFE"

complement(5572..5676)

/locus\_tag="SE1173"

complement(5572..5676)

/locus\_tag="SE1173"

/codon\_start=1

/transl\_table=11

/product="hypothetical protein"

/protein\_id="AAO04772.1"

/db\_xref="GI:27315637"

/translation="MSSLVPNKQPMNHKIOFINGSIHLIADYFIVYG"

complement(5885..6424)

/locus\_tag="SE1174"

complement(5885..6424)

/locus\_tag="SE1174"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAO04773.1"

/db\_xref="GI:27315638"

/translation="MQQKRLLITISLMSAVAVLTIFIKFPIPPYLTLDPSDVPTL

LATLELSPAGIIVALIKNLINLNFNIGDPGVPANFLAGVSLSSYIVYTRKRNKR

SLIYGLITGTIVMTIVLSILNYFVLLPLYGMIFNLGDVLNNVIVISVGVIPENLIG

IIIGIIFVLLFRRLRHIK"

complement(7016..8782)

/locus\_tag="SE1175"

complement(7016..8782)

/locus\_tag="SE1175"

/codon\_start=1

/transl\_table=11

/product="staphylococcal respiratory response protein

SrrB"

/protein\_id="AAO04774.1"

/db\_xref="GI:27315639"

/translation="MNRLNSVYIKMLTILIVATVLVLLSVSLITFIQYFTQETEN

AIKEDAKRISSLVKSHNKTIAIQNSQTLIDGPGGLIIMKMSKDPSSYKDTKKQ

MFNEIKKSTKFKYKEGETQNTITIKNGNSQSYLLGLYPMKAQGAQSHYSGVFI

YKDLKSTEDTNNAITIIITAIITIASTIAPFLSNRIKPLRQLTKQAQKVSQEGD

YSQISTVATKDEIGDLSAFNNMVEIQEHKAISSKNIRDTLLNSMVEGLGNNO

REILSNKMADDIMRHDDPFKSEIEQIEQIEATFESQOQNEYLELEINTRYVFISSYID

RIQNGRSQGIYVMVIRDMTNHNLQMKKDFIANVSHELURTPIQLQHTESLVDGIVT

EPDETIRSLAIVLDESRLNRLNVELNVARMDAEGLSVEKELQPIQLHLLQMSKSYR

MQSELGLTMTFDSNDEQLWNVDMDQVLTNLIDNATRYTQAGDSIKISIDSDSD

NIILITDTGTGIAPEHLKQVDFYKVDAAARKKGQCTGLGLFCIKMIIIEHGGRID

VESLKGKTSFIILPKSKQIS"

complement(8766..9491)

/locus\_tag="SE1176"

complement(8766..9491)

/locus\_tag="SE1176"

/codon\_start=1

/transl\_table=11

/product="staphylococcal respiratory response protein

SrrA"

/protein\_id="AAO04775.1"

/db\_xref="GI:27315640"

/translation="WTNEILIVDDDEDIRRLKLYLRESFEIHEARDKEAYELALE

NNYACILDLMLPEMDGIEVASKREHKDTIIMLTAKGENTNRVEGESGADDYIVK

PFSPREVVLYRKALLRRQTANTQSEPHARDIIEFSLVTDNDARRVLDDQQVNLIT

PKYELVLRVLAKTNNKTRFQEBQQLKVVHYEFYGLDRTVDTHVKKLRKLNKRVSSDA

QMIQTVMGVGYKFEVPSDEPTK"

complement(9625..10362)



```
CDS      /locus_tag="SE1177"
          complement(9625..10362)
          /locus_tag="SE1177"
          /codon_start=1
          /transl_table=11
          /product="ribosomal large subunit pseudouridine synthase
          B"
          /protein_id="AA004776.1"
          /db_xref="GI:27315641"
          /translation="MSKEQLKQRIANSGYTSRRKAENVLITEGRVKNVGQTVTELGS
          KWPMSLVGDKIEQEDKIYLFYKPTQVITSVSDRRVVDYFDLLESRIYFPV
          GRLDYTSVGLLTNDGEFTNLMTHTPRYHKIKKYAKGLGRLMEBEIKQLEKGBLED
          GYTPAQVKKQDKENITLVEITSEGRNQRVRMFYFGHQVKNLTRIQFGPIDL
          KGLNAGEGRVLTPHEVKILRLHLAENGK"
          complement(10355..10897)
          /locus_tag="SE1178"
          /locus_tag="SE1178"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AA004777.1"
          /db_xref="GI:27315642"
          /translation="MDNIAILEALLYTSGDEGLEQKQIIDILDINLQLEDLVSKYHS
          HGLTIQRYGSTVLTATTKETSTVIEQLVKEKSKMLSQAMETLSIIAYNQPLTRGDI
          EMRTGNSDGAVKTLTARGLVEAKVDHRSRSHLLITDPLFNVFGIENLDALPTTEED
          EAEMDEFFSNLVNQKESNE"
          complement(10890..11690)
          /locus_tag="SE1179"
          /locus_tag="SE1179"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AA004778.1"
          /db_xref="GI:27315643"
          /translation="MYSCYNFVKYALTMQFDIEVDMYEVKLDFAFGPLDLLHLIQ
          KYEIDYDIPMKALTSQYQYVHAMQLEINVAEYLVAESLLMIKSKLLLPOTSIE
          EDIEDREDRLVCRLEYQNYKEYTEILNTMKERDLYTKHPTDLTHLETRESWDPN
          QTIDILVELIYAVYQVRKNVELTPKSVIEKTFITIQQAQVTERLKQHESFNFFSL
          FTFPHEVQVVTHTFLAILLEMSKSGIIVNIKQTKQFDIDDIIRGVNYSIG"

Query Match      80.8%; Score 20.2; DB 1; Length 300029;
Best Local Similarity 88.08; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 CAAATTAAGAGACTATTCTGTCAA 25
          |||||
          179897 CAAATTAATGAGAAATTCGTGCAA 179873

RESULT 17
AC134559/c
LOCUS      AC134559.5 180692 bp DNA linear ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-332C4 from chromosome 8, complete
sequence.
ACCESSION AC134559
VERSION   AC134559.5 GI:31746699
KEYWORDS  HTG.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 180692)
AUTHORS   Isak.A.
TITLE     The sequence of Mus musculus BAC clone RP24-332C4
JOURNAL   Unpublished (2001)
REFERENCE 2 (bases 1 to 180692)
AUTHORS   Wilson,R.
TITLE     Sequencing of Mus musculus
JOURNAL   Unpublished (2001)
REFERENCE 3 (bases 1 to 180692)
```

```
AUTHORS   McPherson,J.D. and Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (27-SEP-2002) Genome Sequencing Center, 4444 Forest Park
          Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 180692)
AUTHORS   Wilson,R.K.
TITLE     Direct Submission
JOURNAL   Submitted (31-MAY-2003) Genome Sequencing Center, 4444 Forest Park
          Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 180692)
AUTHORS   Wilson,R.K.
TITLE     Direct Submission
JOURNAL   Submitted (15-JUN-2003) Genome Sequencing Center, 4444 Forest Park
          Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 180692)
AUTHORS   Wilson,R.
TITLE     Direct Submission
JOURNAL   Submitted (27-NOV-2003) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          On Jun 15, 2003 this sequence version replaced gi:31249897.
          ----- Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu
          Contact: submissions@watson.wustl.edu
          ----- Summary Statistics
          -----
          Center project name: M_BB0332C04
          -----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:  
The RPCI-24 BAC Library has been constructed by Pieter de Jong and coworkers (http://www.chori.org) from male C57BL/6J mouse spleen and/or brain genomic DNA. The clone and detailed information can be obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:  
This sequence is the entire insert of the clone. This clone is overlapped by AC125181.

```
FEATURES             Location/Qualifiers
     source            1..180692
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="8"
                     /map="8"
                     /clone="RP24-332C4"
                     /clone_lib="RPCI-24"
     repeat_region     3651..3810
                     /rpt_family="B4"
     repeat_region     8421..8518
                     /rpt_family="Alu"
     repeat_region     9813..9926
                     /rpt_family="B4"
     repeat_region     12341..12502
```

```
repeat_region /rpt_family="B2"
15504. .15570
repeat_region /rpt_family="ERV1"
16825. .17053
repeat_region /rpt_family="B2"
17236. .17289
repeat_region /rpt_family="L1"
17354. .17757
repeat_region /rpt_family="MaLR"
18008. .18168
repeat_region /rpt_family="B4"
19175. .19336
repeat_region /rpt_family="B4"
22885. .23056
repeat_region /rpt_family="B4"
23404. .23749
repeat_region /rpt_family="L1"
23811. .23901
repeat_region /rpt_family="MIR"
23929. .24118
repeat_region /rpt_family="L1"
24400. .24592
repeat_region /rpt_family="B4"
25087. .25192
repeat_region /rpt_family="Alu"
25088. .25233
repeat_region /rpt_family="B4"
28351. .28546
repeat_region /rpt_family="ERVK"
30305. .31233
repeat_region /rpt_family="L1"
31379. .31463
repeat_region /rpt_family="MIR"
32867. .32941
repeat_region /rpt_family="L1"
33285. .33333
repeat_region /rpt_family="Alu"
33334. .33429
repeat_region /rpt_family="L1"
33480. .33671
repeat_region /rpt_family="B2"
33707. .33888
repeat_region /rpt_family="MER1_type"
33878. .34150
repeat_region /rpt_family="MaLR"
34657. .34794
repeat_region /rpt_family="Alu"
34663. .34826
repeat_region /rpt_family="B4"
35656. .35799
repeat_region /rpt_family="Alu"
36988. .37362
repeat_region /rpt_family="MaLR"
37616. .37803
repeat_region /rpt_family="B2"
37970. .38064
repeat_region /rpt_family="Alu"
38235. .38596
repeat_region /rpt_family="L1"
38865. .38987
repeat_region /rpt_family="Alu"
41406. .41521
repeat_region /rpt_family="B4"
42792. .42940
repeat_region /rpt_family="B4"
43582. .44132
repeat_region /rpt_family="L1"
44171. .45313
repeat_region /rpt_family="L1"
45384. .45868
repeat_region /rpt_family="L1"
47036. .47108
repeat_region /rpt_family="ID"

52478. .52602
repeat_region /rpt_family="L1"
52616. .52810
repeat_region /rpt_family="B2"
52898. .53722
repeat_region /rpt_family="L1"
55180. .55382
repeat_region /rpt_family="B2"
55342. .55497
repeat_region /rpt_family="B4"
56728. .56848
repeat_region /rpt_family="L1"
56849. .57044
repeat_region /rpt_family="B2"
57045. .57126
repeat_region /rpt_family="L1"
57325. .57510
repeat_region /rpt_family="B2"
57815. .57929
repeat_region /rpt_family="Alu"
57993. .58231
repeat_region /rpt_family="L1"
58315. .58522
repeat_region /rpt_family="B2"
59688. .59871
repeat_region /rpt_family="B2"
60098. .60283
repeat_region /rpt_family="ERVK"
60321. .60395
repeat_region /rpt_family="ID"
60409. .60535
repeat_region /rpt_family="B4"
61091. .61416
repeat_region /rpt_family="MaLR"

Query Match 79.2%; Score 19.8; DB 9; Length 180692;
Best Local Similarity 91.3%; Pred. No. 3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 AAATTAAAGAGACTATTTCGTGCA 24
Db 125231 AATTAAAGATACTATTTCGTGCA 125209
|||||
|||||

RESULT 18
AC125181/c 223306 bp DNA linear ROD 11-NOV-2003
LOCUS Mus musculus BAC clone RP23-324L6 from chromosome 8, complete
DEFINITION sequence.
ACCESSION AC125181
VERSION AC125181.4 GI:28927902
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 223306)
AUTHORS Szelesten-Shahid,S., Kozlowicz,A., Doebber,A., Shahid,S.,
Haglund,K. and Haakenson,W.
TITLE The sequence of Mus musculus BAC clone RP23-324L6
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 223306)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 223306)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 223306)
AUTHORS McPherson,J.D. and Waterston,R.H.
```



```

/rpt_family="MER1_type"
58372..58931
/rpt_family="ERVK"
58932..59080
/rpt_family="Alu"
59081..59388
/rpt_family="ERVK"
59595..59923
/rpt_family="L1"
60655..61177
/rpt_family="L1"
61370..61844
/rpt_family="L1"
61845..61991
/rpt_family="Alu"
61992..62394
/rpt_family="L1"
62504..62551
/rpt_family="ERV1"
63152..69230
/rpt_family="ID"
70669..70846
/rpt_family="RMR15"
73521..73615
/rpt_family="Alu"
73651..73849
/rpt_family="B2"

```

```

Query Match          79.2%; Score 19.8; DB 9; Length 223306;
Best Local Similarity 91.3%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 2 AAATTAAGAGACTATTCGTGCA 24
|| ||||| ||||| ||||| |||||
Db 2721 AAATTAAGATACACTATTCGTGCA 2699

```

```

RESULT 19
AC113904
LOCUS      233163 bp      DNA      linear      HTG 19-NOV-2002
DEFINITION Rattus norvegicus clone CH230-318P21, *** SEQUENCING IN PROGRESS
*** 2 unrounded clones.
AC113904
AC113904.5 GI:25072640
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.
1 (bases 1 to 233163)
Muzny,D,Marie,, Metzker,M, Lee,, Abramson,S,, Adams,C,, Alder,J.,
Allen,C,, Allen,H,, Alsbrooks,S,, Amin,A,, Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalo,K., Blair,J., Blankenburg,K., Blych,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregregias,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,N., Hamilton,K.,
Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

```

```

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensunewa, L., Louised, H., Lozado, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhiney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newcon, N., Nguyen, N., Norris, S., Parks, K.,
Nwaokemele, O., Okwuonu, G., Olarpunasegon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkoch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ben, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, J., Rojas, A., Rose, M., Rose, R., Ruiz, S.D.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.
Direct Submission
2 (bases 1 to 233163)
Worley, K.C.
Direct Submission
Submitted (05-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 233163)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23194968.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTFV
Center clone name: CH230-318P21
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 209814 bases at least Q40
Consensus quality: 211069 bases at least Q30
Consensus quality: 211999 bases at least Q20
Estimated insert size: 220261; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently

```

\* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 223189: contig of 223189 bp in length  
 \* 223190 223289: gap of unknown length  
 \* 223290 233163: contig of 9874 bp in length.

## FEATURES

source Location/Qualifiers  
 1. 233163

/organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-318P21"

## misc\_feature

1. 2826

/note="wgs\_end\_extension"

## misc\_feature

7180 - 8127

/note="clone\_boundary"

clone\_end:T7

site:

end\_sequence=BZ141951"

223190 .223289

/estimated\_length=unknown

## ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 233163;  
 Best Local Similarity 91.3%; Pred. No. 2.8e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ATTAAGACATTCGTGCA 25

Db 13272 ATTAAGACATTCGTGCA 13294

## RESULT 20

## AC126298/c

LOCUS AC126298 235700 bp DNA linear HTG 22-SEP-2002  
 DEFINITION Rattus norvegicus clone CH230-31B17, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 9 unordered pieces.

AC126298 3 GI-23196133

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 235700)

Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,  
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,  
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
 Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A.,  
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cres,A., D'Souza,L.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
 Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
 Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
 Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J.,  
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
 Hollins,B., Howells,S., Huliyk,S., Hume,J., Idlebird,D., Jackson,A.,  
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowls,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
 Lorusshewa,L., Louieged,H., Lozado,R.J., Lu,X., Ma,J.,  
 Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
 Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,  
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Naif,L.,  
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,  
 Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,  
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,  
 Popper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,  
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,  
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,  
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,  
 Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shan,H.,  
 Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajs,D.,  
 Sneed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J.,  
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,F., Taylor,C.,  
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,  
 Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,  
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,P.,  
 Williams,G., Willson,R., Wlarczyk,R., Wooden,H., Worley,K.,  
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von  
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
 Weinstock,G. and Gibbs,R.A.

## TITLE

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 235700)

AUTHORS Worley,K.C.

TITLE Direct Submission

JOURNAL Submitted (05-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 235700)

AUTHORS Rat Genome Sequencing Consortium.

TITLE Direct Submission

JOURNAL Submitted (22-SEP-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

On Sep 19, 2002 this sequence version replaced gi:21702798.

The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
 sequence may extend beyond the ends of the clone and there may be  
 contigs that consist entirely of whole genome shotgun sequence  
 reads. Both end sequences and whole genome shotgun sequence only  
 contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GYMB

Center Clone name: CH230-31B17

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 161627 bases at least Q40

Consensus quality: 168027 bases at least Q30

Consensus quality: 17221 bases at least Q20

Estimated insert size: 192265; sum-of-contigs estimation

Quality coverage: 3x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 9 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 10630: contig of 10630 bp in length  
\* 10631 10730: gap of unknown length  
\* 10731 22411: contig of 21381 bp in length  
\* 22411 22412: gap of unknown length  
\* 22412 22527: contig of 1064 bp in length  
\* 22527 22537: gap of unknown length  
\* 22537 22638: contig of 1009 bp in length  
\* 22638 22648: gap of unknown length  
\* 22648 22808: contig of 1599 bp in length  
\* 22808 22818: gap of unknown length  
\* 22818 22928: contig of 1045 bp in length  
\* 22928 22938: gap of unknown length  
\* 22938 23072: contig of 1395 bp in length  
\* 23072 23082: gap of unknown length  
\* 23082 23136: contig of 2313 bp in length  
\* 23136 23236: gap of unknown length  
\* 23236 23570: contig of 2464 bp in length.  
\* 23570 23571: contig of 2464 bp in length.

## FEATURES

Source  
1. 235700  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-31B17"

## misc\_feature

1. 1234

/notes="wgs contig"

10631. 10730

/estimated\_length=unknown

## misc\_feature

30002. 31205

/notes="wgs contig"

42447. 43559

/notes="wgs contig"

56213. 57746

/notes="wgs contig"

103185. 104507

/notes="wgs contig"

121152. 121288

/notes="clone boundary"

clone end:T7

site:ECORI

end sequence:BH319296"

220901. 222807

/notes="wgs\_end\_extension"

clone end:T7"

224112. 224211

/estimated\_length=unknown

225276. 225375

/estimated\_length=unknown

226385. 226484

/estimated\_length=unknown

228084. 228183

/estimated\_length=unknown

229229. 229328

/estimated\_length=unknown

230724. 230823

/estimated\_length=unknown

23137. 231236

/estimated\_length=unknown

## ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 235700;  
Best Local Similarity 91.3%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AATTAAAGAGACTATTCGTCAA 25

Db 139577 AATTAAAGAGACTAATTGTGCAA 139555

RESULT 21

AC121708/c

. LOCUS AC121708 237969 bp DNA linear HTG 21-SEP-2002

## DEFINITION

## ACCESSION

AC121708.3

## VERSION

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

Rattus norvegicus clone CH230-190A9, \*\*\* SEQUENCING IN PROGRESS  
\*\*\* 3 unordered pieces.

AC121708

GI:23265128

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Rattus.

1 (bases 1 to 237969)

Muzny, D. Marie., Metzker, M. Lee., Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, P., Garcia, A., Garner, T., Garza, M.,  
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerrero, I., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensu, H., L., Loule, H., Lozado, R. J., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwankwelu, O., Okwuonu, G., Olarnpunaagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.,  
Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajda, D.,  
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,  
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von  
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,  
Weinstock, G. and Gibbs, R. A.

## Direct Submission

Unpublished

2 (bases 1 to 237969)

Worley, K. C.

Direct Submission

Submitted (21-MAY-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 237969)

Rat Genome Sequencing

Direct Submission

Submitted (21-SEP-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

## COMMENT

Baylor Plaza, Houston, TX 77030, USA  
 On Sep 21, 2002 this sequence version replaced gi:21909182.  
 The sequence in this assembly is a combination of BAC based reads  
 and whole genome shotgun sequencing reads assembled using Atlas  
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the  
 sequence may extend beyond the ends of the clone and there may be  
 contigs that consist entirely of whole genome shotgun sequence  
 reads. Both end sequences and whole genome shotgun sequence only  
 contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine  
 Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GYGS

Center clone name: CH230-190A9

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 169159 bases at least Q40

Consensus quality: 174293 bases at least Q30

Consensus quality: 177883 bases at least Q20

Estimated insert size: 191644; sum-of-contigs estimation

Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

-----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
 \* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 233860: contig of 233860 bp in length

\* 233861 233960: gap of unknown length

\* 233961 235605: contig of 1645 bp in length

\* 235606 235705: gap of unknown length

\* 235706 237969: contig of 2264 bp in length.

----- Location/Qualifiers

1. .237969

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-190A9"

1. .1164

/note="wgs end\_extension

clone\_end:T7"

1498. .3143

/note="wgs\_end\_extension

clone\_end:T7"

complement(4467. .5177)

/note="clone\_boundary

clone\_end:T7

site:EcoRI

end\_sequence:BH362335"

68167. .70215

/note="wgs contig"

104654. .105710

/note="wgs contig"

108284. .108038

/note="wgs contig"

200400. .201597

/note="wgs contig"

complement(233121. .233860)

/note="clone\_boundary

clone\_end:Sp6

site:EcoRI

end\_sequence:BH362336"

233861. .233960

/estimated\_length=unknown

235606. .235705

gap

gap

/estimated\_length=unknown

## ORIGIN

Query Match 79.2%; Score 19.8; DB 14; Length 237969;  
 Best Local Similarity 91.3%; Pred. No. 2.8e+02;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 3 AATTAAGAGACTATTCGTGCAA 25  
 |||||  
 Db 46221 AATTAAGAGACTAATTTGTCAA 46199

RESULT 22  
 AC074330/c

LOCUS

DEFINITION

AC074330

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Nov 19, 2000 this sequence version replaced gi:9885992.

----- Genome Center

Center: Genome Therapeutics Corporation

Center code: GTC

Web site: http://www.genomecorp.com/

Contact: gtc-seqcenter@genomecorp.com

----- Project Information

Center project name: mg038

----- Summary Statistics

Sequencing vector: N/A

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 990315

Consensus quality: 223772 bases at least Q40

Consensus quality: 228225 bases at least Q30

Consensus quality: 230837 bases at least Q20

Insert size: 242943; sum-of-contigs

Quality coverage: 5.0x in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 35 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1003: contig of 1003 bp in length

\* 1004 1103: gap of unknown length

\* 1104 2262: contig of 1159 bp in length

\* 2263 2362: gap of unknown length

\* 2363 3424: contig of 1061 bp in length

\* 3424 3523: gap of unknown length

\* 3524 4714: contig of 1191 bp in length

\* 4715 4814: gap of unknown length

\* 4815 5945: contig of 1131 bp in length

\* 5946 6046: gap of unknown length

\* 6046 7182: contig of 1137 bp in length

\* 7183 7282: gap of unknown length

```
* 7283      8431: contig of 1149 bp in length
* 8432      8531: gap of unknown length
* 8532      9679: contig of 1148 bp in length
* 9680      9779: gap of unknown length
* 9780     10902: contig of 1123 bp in length
* 10903     11002: gap of unknown length
* 11003     12271: contig of 1269 bp in length
* 12272     12371: gap of unknown length
* 12372     13579: contig of 1208 bp in length
* 13580     13679: gap of unknown length
* 13680     14974: contig of 1295 bp in length
* 14975     15074: gap of unknown length
* 15075     16120: contig of 1046 bp in length
* 16121     16220: gap of unknown length
* 16221     17554: contig of 1334 bp in length
* 17555     17654: gap of unknown length
* 17655     18714: contig of 1060 bp in length
* 18715     18814: gap of unknown length
* 18815     20113: contig of 1399 bp in length
* 20114     20313: gap of unknown length
* 20314     21457: contig of 1144 bp in length
* 21458     21557: gap of unknown length
* 21558     22702: contig of 1145 bp in length
* 22703     22802: gap of unknown length
* 22803     24679: contig of 1877 bp in length
* 24680     24779: gap of unknown length
* 24780     26082: contig of 1303 bp in length
* 26083     26182: gap of unknown length
* 26183     27553: contig of 1371 bp in length
* 27554     27653: gap of unknown length
* 27654     28960: contig of 1307 bp in length
* 28961     29060: gap of unknown length
* 29061     30604: contig of 1544 bp in length
* 30605     32438: contig of 1734 bp in length
* 32439     32538: gap of unknown length
* 32539     34339: contig of 1701 bp in length
* 34340     34340: gap of unknown length
* 34340     35795: contig of 1456 bp in length
* 35796     35895: gap of unknown length
* 35896     37708: contig of 1813 bp in length
* 37709     37808: gap of unknown length
* 37809     38864: contig of 1056 bp in length
* 38865     38964: gap of unknown length
* 38965     41365: contig of 2401 bp in length
* 41366     41465: gap of unknown length
* 41466     43774: contig of 2309 bp in length
* 43775     43874: gap of unknown length
* 43875     46518: contig of 2644 bp in length
* 46519     46618: gap of unknown length
* 46619     49941: contig of 3323 bp in length
* 49942     50041: gap of unknown length
* 50042     101123: contig of 51082 bp in length
* 101124     101223: gap of unknown length
* 101224     162347: contig of 61124 bp in length
* 162348     162447: gap of unknown length
* 162448     246244: contig of 83797 bp in length.
```

## FEATURES

## source

```
1. .246244
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/chromosome="12"
/clone="RP23-439L22"
/clone_lib="RPCI-23"
1. .1003
/note="assembly_name:Contig4"
1004. .1103
/estimated_length=unknown
1104. .2262
/note="assembly_name:Contig15"
2263. .2362
```

```
misc_feature
2363. .3423
/note="assembly_name:Contig34"
3424. .3523
/estimated_length=unknown
3524. .4714
/note="assembly_name:Contig37"
4715. .4814
/estimated_length=unknown
4815. .5945
/note="assembly_name:Contig42"
5946. .6045
/estimated_length=unknown
6046. .7182
/note="assembly_name:Contig43"
7183. .7282
/estimated_length=unknown
7283. .8431
/note="assembly_name:Contig46"
8432. .8531
/estimated_length=unknown
8532. .9679
/note="assembly_name:Contig47"
9680. .9779
/estimated_length=unknown
9780. .10902
/note="assembly_name:Contig48"
10903. .11002
/estimated_length=unknown
11003. .12271
/note="assembly_name:Contig52"
12272. .12371
/estimated_length=unknown
12372. .13579
/note="assembly_name:Contig53"
13580. .13679
/estimated_length=unknown
13680. .14974
/note="assembly_name:Contig55"
14975. .15074
/estimated_length=unknown
15075. .16120
/note="assembly_name:Contig61"
16121. .16220
/estimated_length=unknown
16221. .17554
/note="assembly_name:Contig65"
17555. .17654
/estimated_length=unknown
17655. .18714
/note="assembly_name:Contig75"
18715. .18814
/estimated_length=unknown
18815. .20213
/note="assembly_name:Contig79"
20214. .20313
/estimated_length=unknown
20314. .21457
/note="assembly_name:Contig81"
21458. .21557
/estimated_length=unknown
21558. .22702
/note="assembly_name:Contig87"
22703. .22802
/estimated_length=unknown
22803. .24679
/note="assembly_name:Contig88"
24680. .24779
```

Query Match 79.2%; Score 19.8; DB 14; Length 246244;  
Best Local Similarity 91.3%; Pred. No. 2.8e+02;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;







AL450063.14:53162..53290,AL450063.14:54999..55097,  
AL450063.14:56062..56144,AL450063.14:67578..67704,  
AL450063.14:68691..68786,AL450063.14:73645..73767,  
AL450063.14:85115..85279,AL450063.14:106142..106267)  
/gene="PRKACB"  
/locus\_tag="RP11-82H13.1-002"  
/standard\_names="OTHUMP0000011663"  
/note="match: proteinins: Sw:P05131 Sw:P05383 Tr:P22694"  
/codon\_start=1  
/product="protein kinase, cAMP-dependent, catalytic, beta"  
/protein\_id="Ca114541.1"  
/db\_xref="GI:55959226"  
/translation="MGNATAKKSEVESVREFLAKAKEDFLKKWNTQNNAGLEDF  
ERKKTGTGSGFRVNLKHKATEQYAMKILDKQVKVLKQIEHTLNKRILIAQVNF  
FLVRLEYAFKNSLYMVEYVPGEMFSLHRIGRFSEPHARFVAAOILVTFYLHS  
LDLIYRDLKPENLIDHOGYIQVTFDFGAKRVKGTWTLCQTPPEYLAPELLISKGYNK  
ADVMWALGLIYEMAGPFFPFADQPIQIYKIVSGKVRFPSPHSSDLKDLRLNLQV  
DLTKRFGNLKNGVSDIKTHKWFATTDWIAIYQKVEAPFIPKFRGSDTSNPDYEE  
DIRVITBKCAKEGGEF"  
17935..18427  
/locus\_tag="RP11-486G15.1-001"  
/pseudo  
17935..18427  
/locus\_tag="RP11-486G15.1-001"  
/note="match: proteins: Sw:P97493 Sw:P97615 Sw:Q95108  
Sw:Q99757 Tr:AA05576 Tr:BAB22037 Tr:BAB27267"  
/pseudo  
/codon\_start=1  
/product="thioredoxin 2 (TXN2) pseudogene"  
61234  
/note="Clone\_left\_end: RP11-82H13"  
ORIGIN  
Query Match 77.6%; Score 19.4; DB 8; Length 63233;  
Best Local Similarity 95.2%; Pred. No. 5.4e+02;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAATTAAAGACTATTCTGTG 22  
Db 20566 AAATTAAAGACTATTCTGTG 20586  
RESULT 25  
AC025340/C  
LOCUS 164485 bp DNA linear HTG 24-MAR-2000  
DEFINITION Homo sapiens chromosome 1 clone RP11-742B6 map 1, WORKING DRAFT  
SEQUENCE, 18 unordered pieces.  
AC025340  
ACCESSION AC025340.2 GI:7321608  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 164485)  
REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Homo sapiens chromosome 1, Clone RP11-742B6  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 164485)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boguski, L., Boukhalil, B., Brown, A., Burkett, G.,  
Campopiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tessaye, S., Theodore, J., Tirrell, A., Traversa, M., Trigilio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 24, 2000 this sequence version replaced gi:7210048.  
All repeats were identified using RepeatMasker:  
Smt, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIER  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L6992  
Center Clone name: 742\_B\_6  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 153825 bases at least Q40  
Consensus quality: 159684 bases at least Q30  
Consensus quality: 161668 bases at least Q20  
Insert size: 152000; agarose-tp  
Insert size: 162785; sum-of-contigs  
Quality coverage: 4.6 in Q20 bases; agarose-tp  
Quality coverage: 4.3 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2576: contig of 2576 bp in length  
\* 2577 2676: gap of 100 bp  
\* 2677 5906: contig of 3230 bp in length  
\* 5907 6006: gap of 100 bp  
\* 6007 9486: contig of 3480 bp in length  
\* 9487 9586: gap of 100 bp  
\* 12061 12160: contig of 2474 bp in length  
\* 12161 18546: contig of 6386 bp in length  
\* 18547 18647: gap of 100 bp  
\* 18647 23010: contig of 4363 bp in length  
\* 23010 29919: contig of 6810 bp in length  
\* 29920 30019: gap of 100 bp  
\* 30020 38748: contig of 8729 bp in length  
\* 38749 38848: gap of 100 bp  
\* 38849 46149: contig of 7301 bp in length  
\* 46150 53642: contig of 7393 bp in length  
\* 53643 53742: gap of 100 bp  
\* 53743 63925: contig of 10183 bp in length  
\* 63926 73519: contig of 9494 bp in length  
\* 73520 73619: gap of 100 bp  
\* 73620 87188: contig of 13569 bp in length  
\* 87189 87288: gap of 100 bp  
\* 87289 100090: contig of 12802 bp in length  
\* 100091 100190: gap of 100 bp  
\* 100191 111124: contig of 10934 bp in length  
\* 111125 111224: gap of 100 bp

TITLE  
JOURNAL

COMMENT

```

* 111225 123063: contig of 11839 bp in length
* 123064 123163: gap of 100 bp
* 123164 141217: contig of 18054 bp in length
* 141218 141317: gap of 100 bp
* 141318 164485: contig of 23168 bp in length.
FEATURES             Location/Qualifiers
     source            1..164485
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="1"
                        /map="1"
     misc_feature       /clone="RP11-742B6"
                        /clone_lib="RPC1-11 Human Male BAC"
                        1..2576
                        /note="assembly_fragment"
     gap                2577..2676
                        /estimated_length=100
     misc_feature       2677..5906
                        /note="assembly_fragment"
     gap                5907..6006
                        /estimated_length=100
     misc_feature       6007..9485
                        /note="assembly_fragment"
     gap                9487..9586
                        /estimated_length=100
     misc_feature       9587..12060
                        /note="assembly_fragment"
     gap                12061..12160
                        /estimated_length=100
     misc_feature       12161..18546
                        /note="assembly_fragment"
     gap                18547..18646
                        /estimated_length=100
     misc_feature       18647..23009
                        /note="assembly_fragment
                        clone_end:17
                        vector_side:left"
     gap                23010..23109
                        /estimated_length=100
     misc_feature       23110..29919
                        /note="assembly_fragment"
     gap                29920..30019
                        /estimated_length=100
     misc_feature       30020..38748
                        /note="assembly_fragment"
     gap                38749..38848
                        /estimated_length=100
     misc_feature       38849..46149
                        /note="assembly_fragment"
     gap                46150..46249
                        /estimated_length=100
     misc_feature       46250..53642
                        /note="assembly_fragment"
     gap                53643..53742
                        /estimated_length=100
     misc_feature       53743..63925
                        /note="assembly_fragment"
     gap                63926..64025
                        /estimated_length=100
     misc_feature       64026..73519
                        /note="assembly_fragment"
     gap                73520..73619
                        /estimated_length=100
     misc_feature       73620..87188
                        /note="assembly_fragment"
     gap                87189..87288
                        /estimated_length=100
     misc_feature       87289..100090
                        /note="assembly_fragment"
     gap                100091..100190
                        /estimated_length=100
     misc_feature       100191..111124

/note="assembly_fragment"
11125..111224
/estimated_length=100
111225..123063
/note="assembly_fragment"
123064..123163
/estimated_length=100
123164..141217
/note="assembly_fragment"
141218..141317
/estimated_length=100
141318..164485
/note="assembly_fragment
clone_end:SP6
vector_side:right"

ORIGIN
Query Match       77.6%; Score 19.4; DB 14; Length 164485;
Best Local Similarity 95.2%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAATTAAGAGACTATTGCTG 22
        |||||
Db      26102 AAATTAAGAGACTATTCTG 26082

RESULT 26
AE007858      11962 bp      DNA      linear      BCT 01-JUN-2004
LOCUS      Clostridium acetobutylicum ATCC 824 section 346 of 356 of the
            complete genome
ACCESSION      AE007858
VERSION      AE007858.1 GI:15026716
KEYWORDS
ORGANISM      Clostridium acetobutylicum ATCC 824
            Clostridium acetobutylicum ATCC 824
            Clostridium acetobutylicum ATCC 824
            Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
            Clostridium.
REFERENCE      1 (bases 1 to 11962)
            Nolling,J., Breton,G., Omelchenko,M.V., Markarova,K.S., Zeng,Q.,
            Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hitti,J., Wolf,Y.I.,
            Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
            Daly,M.J., Bennett G.N., Koonin,E.V. and Smith,D.R.
            Genome sequence and comparative analysis of the solvent-producing
            bacterium Clostridium acetobutylicum
            J. Bacteriol. 183 (16), 4823-4838 (2001)
            11466286
REFERENCE      2 (bases 1 to 11962)
            Childress,D., Zeng,Q. and Smith,D.R.
            Direct Submission
            Submitted (24-JUL-2001) GTC Sequencing Center Production,
            Finishing, and Bioinformatics Teams, Genome Therapeutics Corp., 100
            Beaver Street, Waltham, MA 02453-8443, USA
FEATURES             Location/Qualifiers
     source            1..11962
                        /organism="Clostridium acetobutylicum ATCC 824"
                        /mol_type="genomic DNA"
                        /strain="ATCC 824"
                        /db_xref="ATCC:824"
                        /db_xref="taxon:272562"
                        268..783
                        /gene="CAC3617"
                        268..783
                        /gene="CAC3617"
                        /codon_start=1
                        /transl_table=11
                        /product="Uncharacterized membrane protein, YHAG
                        B. subtilis homolog"
                        /protein_id="AAK81540.1"
                        /db_xref="GI:15026717"
                        /translation="MRMNLKKLIINSLFLAVGVNLNQITPPILFGMKPFDLSAMLFII
                        ILLNDYKTCISTGVVAGLAAAVTTPGGQLPNIDRIVTTSLVFIALRPFKDKIND
                        KIHMTITTVGTIISGSVFLGSAIIIVGLPASFKALFITVVLPAIIINAVGTIIFVA

```

gene	VKSMRVVFRQ" complement (848. .1579) /gene="CAC3618" complement (848. .1579) /gene="CAC3618" /codon_start=1 /transl_table=11 /product="ABC-type polar amino acid transport system, ATPase component" /protein_id="AAK81541.1" /db_xref="GI:15026718" /translation="MLMIAKNLTTFGLKTVFEDLNVNNSISGEVLVIIGPSSGSKST FURCLNHEEPNGGQVIEGEVNHDKKKYREIIEKTMVFQSNLPHMVLQNM EAPITVKENKSEVKRAEALYKVLGSDNNVPSKLSGGQKQVAFARALCMOPDI MLFDPTKSALDELVEVLVSNKELAEQMTWVVTHEMGPAFEVADRVI FWDGGKIV BENTPEGIFTNPKERTKEFLNKIL" complement (1593. .2246) /gene="CAC3619" complement (1593. .2246) /gene="CAC3619" /codon_start=1 /transl_table=11 /product="Amino acid ABC transporter, permease component" /protein_id="AAK81542.1" /db_xref="GI:15026719" /translation="WDISLNKVPVLLDGTTRITLLTCTSSIIIGCIITIAMFKTS SVKVLNLSKFTWILRGTPLLQLYVYVYGLPFLSDKLTMTPKAAIILGLSLNSGAY IARIINGGLAINDGQFEASKALGLTYGQTKRILPQAIRVIVPPCGNEFAMIKDT SLVSVITMBELLRKAQLVSSSGDAVTPYFAGIFVLITTIPTGIFSKIEKKLSVY" complement (2311. .3105) /gene="CAC3620" complement (2311. .3105) /gene="CAC3620" /codon_start=1 /transl_table=11 /product="Amino acid (probably glutamine) ABC transporter, periplasmic binding protein component" /protein_id="AAK81543.1" /db_xref="GI:15026720" /translation="MLKIFLSLVVLTLAIFTGCTSATVSKDLSLRVKAKAIEVI GDDTYPMEFKATGVSDFIDMANAIKAKLGVKTKFVPSFDGIFLALKSKKFD VHSSISITIDERRKAMIFSDPIYGGNAIFVKSDNTTIKNEKDFNNQVVGCVGTAAQ DVLAKISIGIKVKKYKNAWTDALDLQNGRIAAVSDPMVDYNYVTDPGFKKLESSL TOEPIGVAFKEDKSLRDAYNKAIRKELKKGDTMSKLSKKWFGIDVY" 3368. .3556 /gene="CAC3621" /note="fer" 3368. .3556 /gene="CAC3621" /codon_start=1 /transl_table=11 /product="Ferredoxin" /protein_id="AAK81544.1" /db_xref="GI:15026721" /translation="MKAFVDKETCIGCGTCALCPPEIFEMEDDGKAVASDAIBDLK ESQDRAESCVDAINVR" 3842. .4843 /gene="CAC3622" 3842. .4843 /gene="CAC3622" /codon_start=1 /transl_table=11 /product="Possible subunit of benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase" /protein_id="AAK81545.1" /db_xref="GI:15026722" /translation="MKKIGLTTVPVLEVLISAGYTPIDLNIFVSGSKYSKYIDIAEK DGPFSLCIAWKIYCAISNDIKIEIVGHGDCSNTKALTEVLSKKGKIKVPPFSPS SHNINVDQELDFMNLGFVTLNQVEKIRLNKRELGEIDKLTFLDKNKALENH IWQISFSDLDNDLIKTADIADKKREASSSKRLGYIGVPPMTPDIYEFVFN DAKIVNEVOREFAFRSSKAKDIYEQYDITYVDYVYFRIEISQIEERHIDIGIH YTOAFYRAVDQIVIEKLRVPVINIEGRDLNHLDTKRLLEAFDLMLDSKEGIF" 4840. .5610	gene
CDS	/gene="CAC3623" 4840. .5610 /gene="CAC3623" /codon_start=1 /transl_table=11 /product="Activator of 2-hydroxyglutaryl-CoA dehydratase (HSP70-class ATPase domain)" /protein_id="AAK81546.1" /db_xref="GI:15026723" /translation="MKTIGIDLGSREVKLALMENNKKICSTFKISTISFYKYNCSYKNN KIEVDLEKLNLSLISVSTGYGRNNTNLSLPKPINEIKAHVYGVYOTNLKDFVLLD IQGDVRAKLVYEGFIKDLLENKCAASCGRYLENMANVLEIPLDEMSPYENPVKLN STGCVFSSELIKIAEGTIERLCAGVNYSNYNNKMKPLLTAKFMKATLVLTGCVANNY AIKKYLSHDYENIIVPGNPNGAIGCCFYGSKLSYKE" 5617. .6027 /gene="CAC3624" 5617. .6027 /gene="CAC3624" /codon_start=1 /transl_table=11 /product="6-pyruvoyl-tetrahydropterin synthase" /protein_id="AAK81547.1" /db_xref="GI:15026724" /translation="WYILKIEHSPDSAHFLNNGKCAHCHRWVEIEIOSETLID DMIEDFTLKRIKEVLDYDHALLIERKNSLRKUTFDCLIEDGFKIILDPRPTAENF SKFFPEIRKLNIDVRKITVYETPNNSASYKVEV" 6030. .6695 /gene="CAC3625" 6030. .6695 /gene="CAC3625" /codon_start=1 /transl_table=11 /product="Fe-S oxidoreductase of MoaA family" /protein_id="AAK81548.1" /db_xref="GI:15026725" /translation="MNYKVKFVSVINGEGLKSGQLSVFIRFAGNLCNVCYDTKWAN EKDVYLTMTKEBILSYITGVKQVTLTGGEPLQDGIVELLNLSDLTURVEIET NGSVLENFLNFKNAPSFTMDYKLPDSSMENFMKTSNFKLNKKQVIFKPVSSLDKLK KAMDITTEPNLSKKTNIYISPVFGRISPETIVDFMKNKLNLDVTLIQIHKKIWNPNK RGV" 6699. .7286 /gene="CAC3626" /note="mta" 6699. .7286 /gene="CAC3626" /codon_start=1 /transl_table=11 /product="GTP cyclohydrolase I" /protein_id="AAK81549.1" /db_xref="GI:15026726" /translation="MSIDTKAIEKHIGILIALGDDPDREGLKETPKRVAKMWVEEVFK GMEYSNDIDIAENPKTFBEDLKESEDNIVLVKDIEIPIHCEHHLTLMYNMTVAVAYI KMLRIGLISKTARIADWYSRRLQLOERIGKDAIEIWEKVTASKDIAVILKGEHGMCTS RGIKKPGALTITTLKGRFKNDDSLVTKLMALYKA" 7311. .7979 /gene="CAC3627" 7311. .7979 /gene="CAC3627" /codon_start=1 /transl_table=11 /product="pp-loop superfamily ATPase, confers aluminum resistance" /protein_id="AAK81550.1" /db_xref="GI:15026727" /translation="MEISINKEKALVWFGSGDSTTCLFWAKKRYKEVAVVAFSDYNQK HKLEECARDICKHGVHHLDMKLNLAQNSLTRADMKYDEBPKDGPPTFVVG RNLFLSPAFAVAKQGRINNIITGVSQSDFSGYPCDRVFIKSLNVLNLAIDYQFVL ITPLWIDKAEWTKLADLDIVKNETLITCYNGIKSGCGCECPACKLRKNGYVEFK RLYK" complement (8010. .8978) /gene="CAC3628" /note="opp" complement (8010. .8978)	CDS
gene	/gene="CAC3629" complement (8010. .8978) /gene="CAC3629" /codon_start=1 /transl_table=11 /product="ABC-type polar amino acid transport system, ATPase component" /protein_id="AAK81551.1" /db_xref="GI:15026728" /translation="MLMIAKNLTTFGLKTVFEDLNVNNSISGEVLVIIGPSSGSKST FURCLNHEEPNGGQVIEGEVNHDKKKYREIIEKTMVFQSNLPHMVLQNM EAPITVKENKSEVKRAEALYKVLGSDNNVPSKLSGGQKQVAFARALCMOPDI MLFDPTKSALDELVEVLVSNKELAEQMTWVVTHEMGPAFEVADRVI FWDGGKIV BENTPEGIFTNPKERTKEFLNKIL" complement (1593. .2246) /gene="CAC3630" complement (1593. .2246) /gene="CAC3630" /codon_start=1 /transl_table=11 /product="Amino acid ABC transporter, permease component" /protein_id="AAK81552.1" /db_xref="GI:15026729" /translation="WDISLNKVPVLLDGTTRITLLTCTSSIIIGCIITIAMFKTS SVKVLNLSKFTWILRGTPLLQLYVYVYGLPFLSDKLTMTPKAAIILGLSLNSGAY IARIINGGLAINDGQFEASKALGLTYGQTKRILPQAIRVIVPPCGNEFAMIKDT SLVSVITMBELLRKAQLVSSSGDAVTPYFAGIFVLITTIPTGIFSKIEKKLSVY" complement (2311. .3105) /gene="CAC3631" complement (2311. .3105) /gene="CAC3631" /codon_start=1 /transl_table=11 /product="Amino acid (probably glutamine) ABC transporter, periplasmic binding protein component" /protein_id="AAK81553.1" /db_xref="GI:15026730" /translation="MLKIFLSLVVLTLAIFTGCTSATVSKDLSLRVKAKAIEVI GDDTYPMEFKATGVSDFIDMANAIKAKLGVKTKFVPSFDGIFLALKSKKFD VHSSISITIDERRKAMIFSDPIYGGNAIFVKSDNTTIKNEKDFNNQVVGCVGTAAQ DVLAKISIGIKVKKYKNAWTDALDLQNGRIAAVSDPMVDYNYVTDPGFKKLESSL TOEPIGVAFKEDKSLRDAYNKAIRKELKKGDTMSKLSKKWFGIDVY" 3368. .3556 /gene="CAC3632" /note="fer" 3368. .3556 /gene="CAC3632" /codon_start=1 /transl_table=11 /product="Ferredoxin" /protein_id="AAK81554.1" /db_xref="GI:15026731" /translation="MKAFVDKETCIGCGTCALCPPEIFEMEDDGKAVASDAIBDLK ESQDRAESCVDAINVR" 3842. .4843 /gene="CAC3633" 3842. .4843 /gene="CAC3633" /codon_start=1 /transl_table=11 /product="Possible subunit of benzoyl-CoA reductase/2-hydroxyglutaryl-CoA dehydratase" /protein_id="AAK81555.1" /db_xref="GI:15026732" /translation="MKKIGLTTVPVLEVLISAGYTPIDLNIFVSGSKYSKYIDIAEK DGPFSLCIAWKIYCAISNDIKIEIVGHGDCSNTKALTEVLSKKGKIKVPPFSPS SHNINVDQELDFMNLGFVTLNQVEKIRLNKRELGEIDKLTFLDKNKALENH IWQISFSDLDNDLIKTADIADKKREASSSKRLGYIGVPPMTPDIYEFVFN DAKIVNEVOREFAFRSSKAKDIYEQYDITYVDYVYFRIEISQIEERHIDIGIH YTOAFYRAVDQIVIEKLRVPVINIEGRDLNHLDTKRLLEAFDLMLDSKEGIF" 4840. .5610	gene

```

/gene="CAC3628"
/codon_start=1
/transl_table=11
/product="Oligopeptide ABC transporter, ATPase component"
/protein_id="AAK81551.1"
/db_xref="GI:15026728"
/translation="MAEENKSDVLEKLVNKKYFNVGHNAITLKAVDDVSFTIKKGETL
GLVSGSGCKTTCGRVTHGLQATEGEVNVGVNHSURGAKKEFSKKAQMIQDDPY
ACLNPRMTVGDIIAEGIDVHGLYKGEETKKIYEMLDHVLNREHSGRFPHEFSGGQR

Query Match      76.8%; Score 19.2; DB 1; Length 11962;
Best Local Similarity 87.5%; Pred. No. 9 2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGCAA 25
Db 1652 AAATTAAGATAATTCCTGCAA 1675

RESULT 27
AC103751/c
LOCUS      67233 bp      DNA      linear      HTG 30-JUL-2002
DEFINITION Homo sapiens chromosome 15 clone RP11-536E9 map 15, LOW-PASS
SEQUENCE SMPILING.
AC103751
AC103751.2 GI:22004476
VERSION     HTG; HTGS PHASE0.
KEYWORDS    Homo sapiens (human)
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE   1 (bases 1 to 67233)
AUTHORS    Birten,B., Nusbaum,C. and Lander,E.
TITLES     Homo sapiens chromosome 15, clone RP11-536E9
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 67233)
AUTHORS    Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLES     Direct Submission
JOURNAL    Submitted (29-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 67233)
AUTHORS    Birten,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,

```

```

Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 30, 2002 this sequence version replaced gi:17149636.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21702
Center clone name: 536_E_9
-----
* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 647 746: contig of 646 bp in length
* 747 1446: gap of 100 bp
* 1447 1446: contig of 700 bp in length
* 1547 1546: gap of 100 bp
* 2253 2252: contig of 706 bp in length
* 2353 2352: gap of 100 bp
* 3043 3042: contig of 690 bp in length
* 3143 3142: gap of 100 bp
* 3840 3839: contig of 697 bp in length
* 3940 3939: gap of 100 bp
* 4617 4616: contig of 677 bp in length
* 4717 4716: gap of 100 bp
* 5423 5422: contig of 706 bp in length
* 5523 5522: gap of 100 bp
* 6233 6232: contig of 710 bp in length
* 6333 6332: gap of 100 bp
* 7044 7043: contig of 711 bp in length
* 7144 7143: gap of 100 bp
* 7856 7855: contig of 712 bp in length
* 7956 7955: gap of 100 bp
* 8653 8652: contig of 637 bp in length
* 8753 8752: gap of 100 bp
* 9463 9462: contig of 710 bp in length
* 9563 9562: gap of 100 bp
* 10261 10260: contig of 698 bp in length
* 10361 10360: gap of 100 bp
* 10361 1041: contig of 681 bp in length
* 11042 11041: gap of 100 bp
* 11142 11848: contig of 707 bp in length
* 11949 11948: gap of 100 bp
* 12654 12653: contig of 705 bp in length
* 12754 12753: gap of 100 bp
* 13440 13439: contig of 686 bp in length
* 13540 13539: gap of 100 bp
* 14260 14259: contig of 720 bp in length
* 14360 14359: gap of 100 bp
* 15078 15077: contig of 718 bp in length
* 15177 15176: gap of 100 bp

```

TITLE  
JOURNAL

COMMENT

\* 15178 15883: contig of 706 bp in length  
 \* 15884 15983: gap of 100 bp  
 \* 15984 16690: contig of 707 bp in length  
 \* 16690 16790: gap of 100 bp  
 \* 16790 17500: contig of 710 bp in length  
 \* 17500 17600: gap of 100 bp  
 \* 17600 18297: contig of 697 bp in length  
 \* 18297 18398: gap of 100 bp  
 \* 18398 19091: contig of 694 bp in length  
 \* 19091 19191: gap of 100 bp  
 \* 19191 19883: contig of 692 bp in length  
 \* 19883 19984: gap of 100 bp  
 \* 19984 20670: contig of 687 bp in length  
 \* 20670 20770: gap of 100 bp  
 \* 20770 21475: contig of 705 bp in length  
 \* 21475 21575: gap of 100 bp  
 \* 21575 22289: contig of 714 bp in length  
 \* 22289 22390: gap of 100 bp  
 \* 22390 23088: contig of 699 bp in length  
 \* 23088 23189: gap of 100 bp  
 \* 23189 23895: contig of 707 bp in length  
 \* 23895 23956: gap of 100 bp  
 \* 23956 24723: contig of 727 bp in length  
 \* 24723 24823: gap of 100 bp  
 \* 24823 25529: contig of 707 bp in length  
 \* 25529 25629: gap of 100 bp  
 \* 25629 26338: contig of 709 bp in length  
 \* 26338 26438: gap of 100 bp  
 \* 26438 27151: contig of 713 bp in length  
 \* 27151 27251: gap of 100 bp  
 \* 27251 27956: contig of 705 bp in length  
 \* 27956 28056: gap of 100 bp  
 \* 28056 28755: contig of 699 bp in length  
 \* 28755 28855: gap of 100 bp  
 \* 28855 29554: contig of 699 bp in length  
 \* 29554 29654: gap of 100 bp  
 \* 29654 30366: contig of 712 bp in length  
 \* 30366 30466: gap of 100 bp  
 \* 30466 31168: contig of 702 bp in length  
 \* 31168 31268: gap of 100 bp  
 \* 31268 31985: contig of 717 bp in length  
 \* 31985 32085: gap of 100 bp  
 \* 32085 32804: contig of 719 bp in length  
 \* 32804 32904: gap of 100 bp  
 \* 32904 33606: contig of 702 bp in length  
 \* 33606 33708: gap of 100 bp  
 \* 33708 34393: contig of 687 bp in length  
 \* 34393 34493: gap of 100 bp  
 \* 34493 35187: contig of 694 bp in length  
 \* 35187 35287: gap of 100 bp  
 \* 35287 35995: contig of 708 bp in length  
 \* 35995 36095: gap of 100 bp  
 \* 36095 36786: contig of 691 bp in length  
 \* 36786 36886: gap of 100 bp  
 \* 36886 37569: contig of 683 bp in length  
 \* 37569 37669: gap of 100 bp  
 \* 37669 38364: contig of 695 bp in length  
 \* 38364 38464: gap of 100 bp  
 \* 38464 39175: contig of 711 bp in length  
 \* 39175 39275: gap of 100 bp  
 \* 39275 39955: contig of 680 bp in length  
 \* 39955 40055: gap of 100 bp  
 \* 40055 40763: contig of 708 bp in length  
 \* 40763 40864: gap of 100 bp  
 \* 40864 41565: contig of 702 bp in length  
 \* 41565 42381: contig of 716 bp in length  
 \* 42381 42481: gap of 100 bp  
 \* 42481 43191: contig of 710 bp in length  
 \* 43191 43291: gap of 100 bp  
 \* 43291 44002: contig of 711 bp in length  
 \* 44002 44102: gap of 100 bp  
 \* 44102 44808: contig of 706 bp in length  
 \* 44808 44908: gap of 100 bp  
 \* 44908 45619: contig of 711 bp in length

\* 44809 44908: gap of 100 bp  
 \* 44909 45619: contig of 711 bp in length  
 Query Match 76.8%; Score 19.2; DB 14; Length 67233;  
 Best Local Similarity 87.5%; Pred. No. 6.5e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 AAATTAAGAGACTATTCTGCA 25  
 |||||  
 Db 55694 AAATTAAGAGACTATTCTGCA 55671  
 |||||  
 RESULT 28  
 AC164249 125319 bp DNA linear PLN 10-AUG-2005  
 Medicago truncatula clone mth2-20n15, complete sequence.  
 DEFINITION  
 AC164249  
 AC164249.5 GI:72095866  
 VERSION  
 HTG.  
 KEYWORDS  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM  
 Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.  
 REFERENCE 1 (bases 1 to 125319)  
 AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.  
 and Roe B.A.  
 TITLE Medicago truncatula BAC Clone mth2-20n15  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 125319)  
 AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.  
 and Roe B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-JUN-2005) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 REFERENCE 3 (bases 1 to 125319)  
 AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.  
 and Roe B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-JUN-2005) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 REFERENCE 4 (bases 1 to 125319)  
 AUTHORS Lin S., Dixon R., May G., Sumner L., Gonzales B., Cook D., Kim D.  
 and Roe B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-AUG-2005) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 COMMENT On Aug 10, 2005 this sequence version replaced gi:68342142.  
 -----  
 Center: Department Of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR  
 -----  
 FEATURES  
 source Location/Qualifiers  
 1. .125319  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-20n15"  
 /clone\_lib="Medicago truncatula BAC library H2"  
 ORIGIN  
 Query Match 76.8%; Score 19.2; DB 15; Length 125319;  
 Best Local Similarity 87.5%; Pred. No. 5.7e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CAAATTAAGAGACTATTCTGCA 24  
 |||||  
 Db 28482 CAAATTAAGAGACTATTCTGCA 28505  
 |||||



```

RESULT 29
AC149834/c
LOCUS
DEFINITION
AC149834
AC149834
VERSION
KEYWORDS
SOURCE
ORGANISM

Zea mays clone ZMMBB0531A01, *** SEQUENCING IN PROGRESS ***, 7
ordered pieces.
AC149834
AC149834
HTG: HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEPIN.
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 138075)
Birren,B., Nusbaum,C., Lander,E., Butler,E., Wing,R., Bharti,A.K.
and Messing,J.
Zea mays, clone ZMMBB0531A01
Unpublished
2 (bases 1 to 138075)
Birren,B., Nusbaum,C., Lander,E., Butler,E., Wing,R., Bharti,A.K.,
Messing,J., Abouelleil,A., Allen,N., Anderson,M., Anderson,S.,
Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collimore,A.,
Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S.,
Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P.,
FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L.,
Grand-pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P.,
Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Teafaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-JUN-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 138075)
Birren,B., Nusbaum,C., Lander,E., Butler,E., Wing,R., Bharti,A.K.,
Messing,J., Abouelleil,A., Allen,N., Anderson,M., Anderson,S.,
Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collimore,A.,
Cook,A., Cooke,P., Corum,B., DeArellano,K., Diaz,J.S., Dodge,S.,
Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P.,
FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Graham,L.,
Grand-pierre,N., Hafez,N., Hagopian,D., Hagos,B., Hall,J.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Liu,X., Lui,A., Mabbitt,R., MacLean,C., Macdonald,P.,
Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nguyen,T., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Teafaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-JUL-2004) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RW/RepeatMasker.html

```

```

----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@broad.mit.edu
-----
Bharti, AK and Messing, J: The Plant Genome Initiative at
Rutgers, Kamen Institute, Rutgers, The State University of New
Jersey, 190 Frelinghuysen Road, Piscataway, NJ 08854, USA
(http://pgir.rutgers.edu)
Butler, E and Wing, R: Arizona Genomics Institute, Biological
Sciences West, 448A, P.O. Box 210088, University of Arizona,
Tucson, AZ 85721, USA (http://www.genome.arizona.edu)
----- Project Information
Center project name: L30346
Center clone name: 531_A_1
----- Consensus Information
This consensus is derived from a shotgun assembly that has been
manually curated. It is the best representation of the BAC that we
can generate without further laboratory work. The draft assembly
has been edited, and if possible, ends identified by vector as well
as by BAC end sequences, and contigs ordered and oriented. Bases
that are not Ns are either above Q20 or manually edited. This
assembly was performed with Arachne (Genome Res. 2002 12: 177-189;
Genome Res. 2003 13: 91-96). All trace files for this project are
available at the NCBI trace repository
(http://www.ncbi.nlm.nih.gov/Traces/trace.fcgi?). An exact list of
reads used in this assembly are available at
http://www.broad.mit.edu/annotation/plants/maize/randomclones.html.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 28159: contig of 28159 bp in length
* 28160 28259: gap of 100 bp
* 28260 42651: contig of 14392 bp in length
* 42652 42751: gap of 100 bp
* 42752 50861: contig of 8110 bp in length
* 50862 50961: gap of 100 bp
* 50962 59561: contig of 8600 bp in length
* 59562 104318: contig of 44657 bp in length
* 104319 104418: gap of 100 bp
* 104419 113136: contig of 8718 bp in length
* 113137 113236: gap of 100 bp
* 113237 138075: contig of 24839 bp in length.
FEATURES
Location/Qualifiers
1..138075
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBB0531A01"
/clone_lib="CUGI Maize B73 BAC HindIII Library"
28160..28259
/estimated_length=100
42652..42751
/estimated_length=100
50862..50961
/estimated_length=100
59562..59561
/estimated_length=100
104319..104418
/estimated_length=100
113137..113236
/estimated_length=100
ORIGIN

```



Query Match 76.8%; Score 19.2; DB 14; Length 138075;  
 Best Local Similarity 87.5%; Pred. No. 5.6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24  
 |||||  
 Db 14350 CAAATTAAGAGACTATTCGTGCA 14327

RESULT 30  
 AC073928 138943 bp DNA linear PRI 30-APR-2005  
 LOCUS Homo sapiens BAC clone RP11-745P9 from 2, complete sequence.  
 DEFINITION  
 AC073928  
 AC073928  
 AC073928.5 GI:16950374  
 VERSION  
 HTG.  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 138943)  
 Kozlowicz, A., Spalding, L. and Cedroni, M.  
 The sequence of Homo sapiens BAC clone RP11-745P9  
 Unpublished (2001)  
 REFERENCE  
 2 (bases 1 to 138943)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (05-JUL-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE  
 3 (bases 1 to 138943)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (16-NOV-2001) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE  
 4 (bases 1 to 138943)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (03-JAN-2002) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE  
 5 (bases 1 to 138943)  
 Waterston, R.  
 Direct Submission  
 Submitted (09-JAN-2002) Department of Genetics, Washington  
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE  
 6 (bases 1 to 138943)  
 Wilson, R.K.  
 Direct Submission  
 Submitted (30-APR-2005) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 COMMENT  
 On Nov 16, 2001 this sequence version replaced gi:16756379.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu  
 Contact: submissions@wustl.edu  
 ----- Summary Statistics  
 -----  
 Center project name: H\_NH0745P09  
 -----

## NOTICE:

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate  
 chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by sequence  
 from more than one subclone; and the assembly was confirmed by

## restriction digest.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,  
 Department of Genetics, Washington University, St. Louis MO. For  
 additional information about the map position of this sequence, see  
 http://genome.wustl.edu

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male  
 donor, as described by Oosagawa, K., Woon, P.Y., Zhao, B., Frengen, E.,  
 Tateno, M., Catanesse, J.J. and de Jong, P.J. (1998) An improved  
 approach for construction of bacterial artificial chromosome  
 libraries. Genomics 51:1-8. The clone may be obtained either from  
 Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong  
 and coworkers at http://www.chori.org  
 VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP11-326J14, 2000 bp overlap.  
 Actual start of this clone is at base position 1 of RP11-745P9;  
 Actual end is at base position 48928 of RP11-326J14.

Single plasmid regions exist between 6749 and 67933; 82507 and  
 82581.

## FEATURES

## source

## Location/Qualifiers

1..138943  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /clone="RP11-745P9"  
 /clone\_lib="RPCI-11"  
 10825..11105  
 /note="CpG island (%GC=61.9, o/e=0.79, #CpGs=22)"  
 misc\_feature  
 75901..76274  
 /note="CpG island (%GC=74.3, o/e=0.54, #CpGs=33)"  
 misc\_feature

## ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 138943;  
 Best Local Similarity 87.5%; Pred. No. 5.6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25  
 |||||  
 Db 3560 AAATTAAGAGACTATTCGTGCAA 3583

## RESULT 31

## AC012362/c

LOCUS Homo sapiens BAC clone RP11-404D15 from 2, complete sequence.  
 DEFINITION  
 AC012362  
 AC012362  
 AC012362.7 GI:16306515  
 VERSION  
 HTG.  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM

## REFERENCE

## AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.  
 1 (bases 1 to 156618)  
 Cotton, M., Maupin, R., Phillips, A., Elliott, G., Boyer, E. and  
 Paulson, E.  
 The sequence of Homo sapiens BAC clone RP11-404D15  
 Unpublished (2001)  
 REFERENCE  
 2 (bases 1 to 156618)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (25-OCT-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 REFERENCE  
 3 (bases 1 to 156618)  
 Waterston, R.H.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS



REFERENCE  
AUTHORS  
TITLE  
JOURNAL

University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
2 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (26-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (27-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (28-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (29-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
6 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (30-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
7 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (31-MAY-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
8 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (05-JUN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
9 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (07-JUN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
10 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (12-JUN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
11 (bases 1 to 176602)  
Waterston,R.  
Direct Submission  
Submitted (07-OCT-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
12 (bases 1 to 176602)  
Waterston,R.H.  
Direct Submission  
Submitted (25-MAR-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
13 (bases 1 to 176602)  
Waterston,R.  
Direct Submission  
Submitted (29-OCT-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
14 (bases 1 to 176602)

AUTHORS  
TITLE  
JOURNAL

Wilson,R.K.  
Direct Submission  
Submitted (05-APR-2005) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Oct 7, 2000 this sequence version replaced gi:8099093.  
FEATURES  
source  
1. .176602  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-290K24"  
/clone\_lib="RPC1-11"  
142050..142278  
misc\_feature  
/note="CpG island (%GC=60.3, o/e=0.91, #CpGs=20)"  
146574..146859  
misc\_feature  
/note="CpG island (%GC=61.2, o/e=0.70, #CpGs=21)"  
ORIGIN  
Query Match 76.8%; Score 19.2; DB 8; Length 176602;  
Best Local Similarity 87.5%; Pred.No.5.3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAATTAAGAGACTATTCGTGCAA 25  
|||||  
DB 96881 AAATTAAGAGACTATTCGTGCTA 96904  
|||||  
RESULT 33  
AC154981  
LOCUS  
DEFINITION  
AC154981  
VERSION  
AC154981.2 GI:68265074  
HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS  
Bos taurus (cow)  
SOURCE  
ORGANISM  
Bos taurus  
182536 bp DNA linear HTG 01-JUL-2005  
Bos taurus clone CH240-27D11, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 12  
unordered pieces.  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
REFERENCE  
1 (bases 1 to 182536)  
Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Adams,C., Alder,J.,  
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angilano,D.,  
Anyalebech,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,  
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,  
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,  
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,  
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,  
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,  
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,  
Daviia,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,  
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,  
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,  
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,  
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,  
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,  
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,  
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,  
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,  
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,  
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,  
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,  
Karpathy,S., Kelly,S., Khan,Z., King,L., Kovar,C.,  
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,  
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,  
Lorensuewa,L., Loulsegod,H., Lozado,R.J., Lu,X., Ma,J.,  
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,  
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,  
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,  
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,

Morgan, M., Morris, K., Morris, S., Munidaea, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., NewCon, N., Nguyen, N., Norris, S., Parks, K., Nwackelme, O., Okwuonu, G., Olarnpunaagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanasa, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, J., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

# TITLE JOURNAL REFERENCE

Unpublished  
2 (bases 1 to 182536)

# AUTHORS JOURNAL

Worley K.C.

# TITLE JOURNAL

Submitted (06-JAN-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

# REFERENCE AUTHORS TITLE JOURNAL

3 (bases 1 to 182536)  
Cow Genome Sequencing Consortium.

# TITLE JOURNAL

# COMMENT

Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Jun 28, 2005 this sequence version replaced gi:57164456.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: FB22

Center clone name: CH240-27D11

----- Summary Statistics

Assembly program: Atlas 3.0;

Consensus quality: 178361 bases at least Q40

Consensus quality: 179461 bases at least Q30

Consensus quality: 180552 bases at least Q20

Estimated insert size: 182232; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 12 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 68663: contig of 68663 bp in length  
\* 68664 68713: gap of 50 bp  
\* 68714 105062: contig of 36349 bp in length  
\* 105063 105555: gap of 493 bp  
\* 105556 137308: contig of 31753 bp in length  
\* 137309 137358: gap of 50 bp  
\* 137359 140589: contig of 3231 bp in length  
\* 140590 140639: gap of 50 bp  
\* 140640 14837: contig of 4198 bp in length  
\* 14838 14887: gap of 50 bp  
\* 14888 147571: contig of 2684 bp in length  
\* 147572 147621: gap of 50 bp  
\* 147622 169515: contig of 21894 bp in length  
\* 169516 169565: gap of 50 bp  
\* 169566 173031: contig of 3466 bp in length  
\* 173032 173131: gap of unknown length  
\* 173132 174955: contig of 1824 bp in length  
\* 174956 175005: gap of 50 bp  
\* 175006 179124: contig of 4119 bp in length  
\* 179125 179224: gap of unknown length  
\* 179225 180668: contig of 1444 bp in length  
\* 180669 180768: gap of unknown length  
\* 180769 182536: contig of 1768 bp in length.

# FEATURES

Location/Qualifiers

1. 182536  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-27D11"  
68664..68713  
/estimated\_length=50  
105063..105555  
/estimated\_length=493  
137309..137358  
/estimated\_length=50  
140590..140639  
/estimated\_length=50  
144838..144887  
/estimated\_length=50  
147572..147621  
/estimated\_length=50  
169516..169565  
/estimated\_length=50  
173032..173131  
/estimated\_length=unknown  
174956..175005  
/estimated\_length=50  
179125..179224  
/estimated\_length=unknown  
180669..180768  
/estimated\_length=unknown

# ORIGIN

Query Match 76.8%; Score 19.2; DB 14; Length 182536;  
Best Local Similarity 87.5%; Pred. No. 5,3e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCTGCAA 25

Db 25615 AAAATGAAGAGACTATTCTGCAA 25638

# RESULT 34

AC146141

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AC146141

Pan troglodytes

HTG.

GI:35073670

Pan troglodytes (chimpanzee)

Pan troglodytes

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

AC146141 182547 bp DNA linear PRI 29-OCT-2003  
Pan troglodytes BAC clone RP43-3A23 from 7, complete sequence.



80 - 89 : 52232  
 90 - 99 : 79090  
 -----  
 Percentage of bases with a quality value >= 40 : 99 %

FEATURES  
source

Location/Qualifiers  
 1..183491  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="14"  
 /clone="R-561B11"  
 /clone\_lib="RPCI-11"  
 17109..17210  
 /note="matching EMBL:X59417  
 RHdb:RH79927  
 dbSTS:STS56350  
 Identified using the e-PCR software (G. Schuler)"  
 101390..101659  
 /note="matching EMBL:M69043  
 RHdb:RH69167  
 dbSTS:STS49057  
 Identified using the e-PCR software (G. Schuler)"  
 101471..101667  
 /note="matching EMBL:C13563  
 RHdb:RH7865  
 dbSTS:STS13729  
 Identified using the e-PCR software (G. Schuler)"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 8; Length 183491;  
 Best Local Similarity 87.5%; Pred. No. 5.3e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CAAATTAAAGAGACTATTCTGCA 24  
 |||||  
 Db 146159 CAAATTAAAGAGACTTTTCATCA 146136

RESULT 36  
 AC157660  
 LOCUS AC157660.2 GI:60543592  
 DEFINITION Mus musculus BAC clone RP23-263116 from chromosome 1, complete sequence.  
 AC157660  
 VERSION AC157660.2 GI:60543592  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 1 (bases 1 to 192026)  
 Hlatka, M. and Kozlowski, A.  
 The sequence of Mus musculus BAC clone RP23-263116  
 Unpublished (2001)  
 2 (bases 1 to 192026)  
 Wilson, R.K.  
 Direct Submission  
 Submitted (24-FEB-2005) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 3 (bases 1 to 192026)  
 Wilson, R.K.  
 Direct Submission  
 Submitted (05-MAR-2005) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 4 (bases 1 to 192026)  
 Wilson, R.K.  
 Direct Submission  
 Submitted (23-APR-2005) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 COMMENT  
 On Mar 5, 2005 this sequence version replaced gi:60223388.  
 ----- Genome Center

Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu>  
 Contact: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
 ----- Summary Statistics  
 -----  
 Center project name: M\_BA0263116  
 -----

## NOTICE:

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e. phred quality >=30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone, fosmid clone or direct clone walk sequence. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to obtain the consensus sequence. The assembly was confirmed by restriction digest.  
 This finishing standard has slightly changed from the previous Human standard. Specifically, standards for regions of low sequence complexity (such as dinucleotide repeats and small unit tandem repeats) have been relaxed. These regions are very prevalent in the mouse genome, and the return on extended finishing efforts is minimal.  
 If a sequence meets the criteria of the above statement, it needs no comments or tags. If the criteria are not met, such as ambiguous bases, then the region is duly annotated.

## MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The BAC Library has been constructed by Kazutoyo Oseawa and Minako Tateo in the laboratory of Pieter de Jong (http://www.chori.org) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

This sequence is the entire insert of the clone.

FEATURES  
source

Location/Qualifiers  
 1..192026  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="1"  
 /clone="RP23-263116"  
 /clone\_lib="RPCI-23"  
 1..13  
 /note="Sequence derived from PCR product of project DNA"  
 282..434  
 /note="Sequence derived from PCR product of project DNA"  
 21398..21399  
 /note="Bacterial transposon insertion in clone excised here"  
 123614..123676  
 /note="Sequence derived from one plasmid subclone."  
 158282..158374  
 /note="Unresolved simple sequence repeat."  
 171648..171695  
 /note="Sequence derived from one plasmid subclone."

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 192026;  
 Best Local Similarity 87.5%; Pred. No. 5.3e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 AAATTAAAGAGACTATTCTGCA 25  
 |||||  
 |||||

```
Db 102489 AATAAAGGAGACTATTAGTGCA 102512

RESULT 37
CNS01RGW/c 194173 bp DNA linear PRI 04-MAY-2001
LOCUS Human chromosome 14 DNA sequence BAC R-355C3 of library RPCI-11
DEFINITION from chromosome 14 of Homo sapiens (Human), complete sequence.
ACCESSION AL160231
VERSION AL160231.4 GI:13992187
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 194173)
Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,
Brottier, P., Cattolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
Levy, M., Eckenberg, R., Brule, T., deBerardinis, V., Cruaud, C.,
Gyapay, G., Saurin, W. and Weissbach, J.
Sequencing of the human chromosome 14
Unpublished
2 (bases 1 to 194173)
Genoscope.
Direct Submission
Submitted (04-MAY-2001) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On May 8, 2001 this sequence version replaced gi:13016588.
----- Genom Center
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: Seqref@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-561B11
Downstream BAC (overlapping the SP6 end) : R-102C24 (AC=AL137818)
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 7.06x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 :
20 - 29 :
30 - 39 : 9
40 - 49 : 115
50 - 59 : 2501
60 - 69 : 7708
70 - 79 : 8382
80 - 89 : 18066
90 - 99 : 49946
107446

Percentage of bases with a quality value >= 40 : 99 %.
-----
FEATURES
source
1. 194173
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-355C3"
/clone_lib="RPCI-11"
72179..72308
/notes="matching EMBL:G07705
RHdb:RH32619
dbSTS:STG32292
Identified using the e-PCR software (G. Schuler)"
95324..95487

STG
STG

/note="matching EMBL:Z39869
RHdb:RHS3631
dbSTS:STS17739
Identified using the e-PCR software (G. Schuler)"

ORIGIN
Query Match 76.8%; Score 19.2; DB 8; Length 194173;
Best Local Similarity 87.5%; Pred. No. 5.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAGGAGACTATTGTCGA 24
Db 3237 CAATTAAGGAGACTTTTCATTC 3214

RESULT 38
AC136716 259603 bp DNA linear ROD 18-SEP-2003
LOCUS Mus musculus, clone RP23-42E21, complete sequence.
DEFINITION AC136716
ACCESSION AC136716
VERSION AC136716.12 GI:34849981
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Murioidea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 259603)
Birren, B., Nusbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 259603)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Bouckhagalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faros, S., Ferreira, P., FitzGerald, M., Gage, D., Galeagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kella, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mhova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (07-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 259603)
Birren, B., Nusbaum, C. and Lander, E.
Direct Submission
Submitted (23-AUG-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 259603)
Birren, B., Nusbaum, C. and Lander, E.
Direct Submission
Submitted (18-SEP-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 18, 2003 this sequence version replaced gi:34146915.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genom Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
```



		Center project name: L26291
		Center clone name: 42_E_21
		-----
FEATURES	Location/Qualifiers	
source	1..259603 /organism="Mus musculus" /mol_type="genomic DNA" /db_xref="taxon:10090" /clone="RP23-42E21" /clone.lib="RPCI-23 Female Mouse BAC"	
misc_feature	1..17744 /note="wgs end extension clone end:SP6"	
repeat_region	2..122 /rpt_family="(CAA)n"	
repeat_region	131..281 /rpt_family="Lx2B"	
repeat_region	809..910 /rpt_family="Lx6"	
repeat_region	1339..1450 /rpt_family="B3"	
repeat_region	2167..2249 /rpt_family="(TCTA)n"	
repeat_region	3644..4463 /rpt_family="LI_MM"	
repeat_region	4601..4685 /rpt_family="Lx2"	
repeat_region	complement(4688..5011) /rpt_family="ORRIA2"	
repeat_region	complement(5013..6744) /rpt_family="ORRIA-int"	
repeat_region	complement(6745..6804) /rpt_family="ORRIA1"	
repeat_region	complement(6811..6876) /rpt_family="ORRIA2"	
repeat_region	6877..9072 /rpt_family="Lx2"	
repeat_region	9080..9272 /rpt_family="Lx4"	
repeat_region	10696..10806 /rpt_family="(TC)n"	
unsure	complement(10709..10744) /notes="single clone coverage"	
unsure	complement(10749) /notes="Probably C, possibly G"	
unsure	complement(11119..11123) /notes"<30 qual SNGL region"	
unsure	complement(11173..11178) /notes"<30 qual SNGL region"	
unsure	complement(11179..11208) /notes="single clone coverage"	
unsure	complement(11179..11187) /notes"<30 qual single clone coverage"	
unsure	complement(11190..11196) /notes"<30 qual single clone coverage"	
repeat_region	11222..11266 /rpt_family="(TC)n"	
repeat_region	11266..11315 /rpt_family="(CA)n"	
repeat_region	14509..14561 /rpt_family="(TATATG)n"	
repeat_region	complement(14638..14722) /rpt_family="Lx9"	
repeat_region	14873..14909 /rpt_family="(TTTA)n"	
repeat_region	15765..15910 /rpt_family="B1_MM"	
repeat_region	16755..16800 /rpt_family="(TC)n"	
repeat_region	17245..17578 /rpt_family="IAPUTRI_MM"	
repeat_region	17630..20231 /rpt_family="IAPUTRI MM-int"	

misc_feature	17745. .17750 /notes="clone_boundary clone end:SP6 site:EcoRI"
repeat_region	20214. .21293 /rpt_family="IAPLTR1_MM-int"
repeat_region	21296. .21620 /rpt_family="IAPLTR1_MM"
repeat_region	22530. .22552 /rpt_family="AT_rich"
repeat_region	22708. .28705 /rpt_family="L1_MM"
repeat_region	28833. .28988 /rpt_family="L3"
repeat_region	29103. .30070 /rpt_family="Lx2B"
repeat_region	30071. .30091 /rpt_family="AT_rich"
repeat_region	30134. .30325 /rpt_family="L3"
repeat_region	30521. .30628 /rpt_family="PB1D7"
repeat_region	30829. .30679 /rpt_family="(TG)n"
repeat_region	30852. .30852 complement(30686. .30852)
repeat_region	30933. .31315 /rpt_family="RMER20"
repeat_region	32012. .32039 /rpt_family="AT_rich"
repeat_region	33135. .33560 /rpt_family="Lx"
repeat_region	34911. .34939 /rpt_family="AT_rich"
repeat_region	35684. .35715 /rpt_family="(CAA)n"
repeat_region	35881. .36058 /rpt_family="(TTTC)n"
repeat_region	37161. .37329 /rpt_family="(TG)n"
repeat_region	37332. .37561 /rpt_family="L1M4"
repeat_region	37614. .37809 complement(37614. .37809)
repeat_region	38093. .38292 complement(38093. .38292)
repeat_region	38695. .38935 /rpt_family="L1M4c"
repeat_region	39316. .39346 /rpt_family="L1M4"
repeat_region	39459. .39721 /rpt_family="AT_rich"
repeat_region	39730. .39847 /rpt_family="Lx"
repeat_region	39848. .40192 /rpt_family="L1M4"
repeat_region	40193. .40245 /rpt_family="ORR1A2"
repeat_region	40247. .40287 /rpt_family="L1M4"
repeat_region	40443. .40487 /rpt_family="(TAA)n"
repeat_region	40572. .40618 /rpt_family="AT_rich"
repeat_region	40631. .40915 /rpt_family="L1M4"
repeat_region	41963. .42127 /rpt_family="(TCTA)n"
repeat_region	42248. .42468 /rpt_family="RLTR11A"
repeat_region	43019. .43166

Query Match

76.8%; Score 19.2; DB 9; Length 259603;



Best Local Similarity 87.5%; Pred. No. 4.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGCTGCA 24  
||||| ||||| ||||| ||||| |||||

Db 150487 CAAATCAAGAGATTATTCGAGCA 150510

RESULT 39  
AC152434/c  
LOCUS AC152434 309502 bp DNA linear HTG 01-JUL-2005  
DEFINITION Bos taurus clone CH240-5P17, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 59  
unordered pieces.

AC152434 3 GI:68227043  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Bos taurus (cow)  
SOURCE

ORGANISM Bos taurus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 309502)

AC152434 309502 bp DNA linear HTG 01-JUL-2005  
unordered pieces.  
AC152434 3 GI:68227043  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
KEYWORDS Bos taurus (cow)  
SOURCE  
ORGANISM Bos taurus  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;  
Pecora; Bovidae; Bovinae; Bos.  
1 (bases 1 to 309502)  
Muzny, D. Marie, Metzger, M. Lee, Abramson, S., Adams, C., Alder, J.,  
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,  
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,  
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,  
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,  
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,  
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,  
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,  
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,  
Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,  
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,  
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,  
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,  
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,  
Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,  
Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,  
Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,  
Harvey, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J.,  
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,  
Hollins, B., Howells, S., Hulyak, S., Hume, J., Idlebird, D., Jackson, A.,  
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,  
Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,  
Kowib, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,  
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,  
Lorensu, H., Lu, J., Lu, X., Lu, X., Lu, X., Ma, J.,  
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,  
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,  
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,  
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,  
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,  
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,  
Nwakoleneh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,  
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,  
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L.,  
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,  
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,  
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,  
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,  
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajic, D.,  
Sneed, A., Sodergren, E., Song, X. Z., Sorelle, R., Sosa, J.,  
Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,  
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,  
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,  
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,  
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,  
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,  
Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von  
Weinstock, G., and Gibbs, R. A.  
Direct Submission  
Unpublished

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

COMMENT

2 (bases 1 to 309502)  
Worley, K.C.  
Direct Submission  
Submitted (08-NOV-2004) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 309502)  
Cow Genome Sequencing Consortium.  
Direct Submission  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jun 26, 2005 this sequence version replaced gi:58037939.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: FAKM  
Center clone name: CH240-5P17  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 285635 bases at least Q40  
Consensus quality: 290153 bases at least Q30  
Consensus quality: 294335 bases at least Q20  
Estimated insert size: 285852; sum-of-contigs estimation  
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html))  
\* NOTE: This sequence may represent more than one clone.  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 59 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
\* 11017: contig of 11017 bp in length  
\* 11018: gap of 612 bp  
\* 11629: contig of 3854 bp in length  
\* 15483: contig of 181 bp  
\* 15484: gap of 181 bp  
\* 15664: gap of 181 bp  
\* 25291: contig of 9627 bp in length  
\* 25292: gap of 50 bp  
\* 25341: gap of 50 bp  
\* 29196: contig of 3855 bp in length  
\* 29246: gap of 50 bp  
\* 29247: contig of 1945 bp in length  
\* 31191: contig of 50 bp  
\* 31192: gap of 50 bp  
\* 31241: gap of 50 bp  
\* 40043: contig of 8802 bp in length  
\* 40044: gap of 50 bp  
\* 40094: gap of 50 bp  
\* 46637: contig of 6544 bp in length  
\* 46688: gap of 50 bp  
\* 58397: contig of 11709 bp in length  
\* 58446: gap of 50 bp  
\* 58447: gap of 50 bp  
\* 63178: contig of 4732 bp in length  
\* 63179: gap of 50 bp  
\* 70691: contig of 7463 bp in length  
\* 70692: gap of unknown length  
\* 74795: contig of 4004 bp in length

TITLE  
JOURNAL

\* 74796 74845: gap of 50 bp  
\* 74846 80620: contig of 5775 bp in length  
\* 80621 80670: gap of 50 bp  
\* 80671 86253: contig of 5583 bp in length  
\* 86254 86303: gap of 50 bp  
\* 86304 86588: contig of 10355 bp in length  
\* 86589 96758: gap of unknown length  
\* 96759 97822: contig of 1064 bp in length  
\* 97823 97921: gap of 99 bp  
\* 97922 104351: contig of 6430 bp in length  
\* 104352 104401: gap of 50 bp  
\* 104402 115482: contig of 11081 bp in length  
\* 115483 115532: gap of 50 bp  
\* 115533 128633: contig of 13101 bp in length  
\* 128634 128733: gap of unknown length  
\* 128734 133221: contig of 4488 bp in length  
\* 133222 133271: gap of 50 bp  
\* 133272 139579: contig of 6308 bp in length  
\* 139580 139629: gap of 50 bp  
\* 139630 148562: contig of 8933 bp in length  
\* 148563 148805: gap of 243 bp  
\* 148806 161450: contig of 12645 bp in length  
\* 161451 161500: gap of 50 bp  
\* 161501 165216: contig of 3716 bp in length  
\* 165217 165266: gap of 50 bp  
\* 165267 167490: contig of 2224 bp in length  
\* 167491 168006: gap of 516 bp  
\* 168007 172330: contig of 4324 bp in length  
\* 172331 173367: gap of 1037 bp  
\* 173368 176298: contig of 2931 bp in length  
\* 176299 176892: gap of 594 bp  
\* 176893 184394: contig of 7492 bp in length  
\* 184395 184434: gap of 50 bp  
\* 184435 192053: contig of 7619 bp in length  
\* 192054 192103: gap of 50 bp  
\* 192104 193660: contig of 1557 bp in length  
\* 193661 195706: gap of 2046 bp  
\* 195707 202230: contig of 6524 bp in length  
\* 202231 202280: gap of 50 bp  
\* 202281 211945: contig of 9665 bp in length  
\* 211946 211995: gap of 50 bp  
\* 211996 232014: contig of 20019 bp in length  
\* 232015 232064: gap of 50 bp  
\* 232065 234686: contig of 2622 bp in length  
\* 234687 234786: gap of unknown length  
\* 234787 237743: contig of 2957 bp in length  
\* 237744 238139: gap of 396 bp  
\* 238140 239574: contig of 1435 bp in length  
\* 239575 239624: gap of 50 bp  
\* 239625 246849: contig of 7225 bp in length  
\* 246850 250034: contig of 3135 bp in length  
\* 250035 250134: gap of unknown length  
\* 250135 251145: contig of 1011 bp in length  
\* 251146 252472: gap of unknown length  
\* 252473 252572: gap of 1227 bp in length  
\* 252573 254244: contig of 1672 bp in length  
\* 254245 254344: gap of unknown length  
\* 254345 255355: contig of 1011 bp in length  
\* 255356 255455: gap of unknown length  
\* 255456 256842: contig of 1387 bp in length  
\* 256843 256942: gap of unknown length  
\* 256943 258008: contig of 1066 bp in length

Query Match 76.8%; Score 19.2; DB 14; Length 309502;  
Best Local Similarity 87.5%; Pred. No. 4.8e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGCA 24  
|||||  
DB 34653 CAAATGAAGAGACTATTAGTGA 34630

RESULT 40  
BX901880\_2/c  
WPCOMMENT

Sequence split into 4 fragments LOCUS BX901880 Accession BX901880

Fragment Name	Begin	End
BX901880_0	1	110000
BX901880_1	100001	210000
BX901880_2	200001	310000
BX901880_3	300001	370132

Continuation (3 of 4) of BX901880 from base 200001 (BX901880 Danio rerio clone DKBY-109,

Query Match 75.2%; Score 18.8; DB 14; Length 110000;  
Best Local Similarity 90.9%; Pred. No. 8.6e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCTGTCGA 25  
|||||  
DB 91179 ATTAAGAGAGACTATTCTGTCGA 91158

RESULT 41

BX323460/c

LOCUS

DEFINITION

Zebrafish DNA sequence from clone DK5YP-47G11 in linkage group 3,

complete sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

HTG.

Danio rerio (zebrafish)

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Ostariophysi;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 202017)

Pelam. S.

Direct Submission

Submitted (27-APR-2004) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Apr 27, 2004 this sequence version replaced gi:46559267.

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: zfish-help@sanger.ac.uk

-----

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all

regions were either double-stranded or sequenced with an alternate

chemistry or covered by high quality data (i.e., phred quality >

30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the

assembly was confirmed by restriction digest, except on the rare

occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information

on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Clone-derived

Zebrafish pUC subclones occasionally display inconsistency over the

length of mononucleotide A/T runs and conserved TA repeats. Where

this is found the longest good quality representation will be

submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat

discovery system (Zhirong Bao and Sean Eddy, submitted), and those

beginning 'drr' were identified by Rick Waterman (Stephen Johnson

lab, WashU). For further information see  
[http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml) DKEYP-47G11  
 is from a Zebrafish BAC library  
 VECTOR: pIndigoBAC-5.

FEATURES source  
 1. .202017  
 Location/Qualifiers  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEYP-47G11"  
 /clone\_lib="DanioKeyypilot"

## ORIGIN

Query Match 75.2%; Score 18.8; DB 5; Length 202017;  
 Best Local Similarity 90.9%; Pred. No. 7.6e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTTCGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 19306 AATTAAGAGACTATTTCGTGCA 19285

## RESULT 42

CR391929/c  
 LOCUS 223724 bp DNA linear VRT 23-FEB-2005  
 DEFINITION Zebrafish DNA sequence from clone DKEYP-86B5, complete sequence.

ACCESSION CR391929  
 VERSION CR391929.9 GI:60219381  
 KEYWORDS HTG.  
 SOURCE Danio rerio (zebrafish)

## ORGANISM

Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 223724)

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (22-FEB-2005) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Feb 23, 2005 this sequence version replaced gi:59891260.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: zfish-help@sanger.ac.uk  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at

[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Clone-derived  
 Zebrafish pUC subclones occasionally display inconsistency over the  
 length of mononucleotide A/T runs and conserved TA repeats. Where  
 this is found the longest good quality representation will be  
 submitted.

Repeat names beginning 'Dr' were identified by the Recon repeat  
 discovery system (Zhitong Bao and Sean Eddy, submitted), and those  
 beginning 'drr' were identified by Rick Waterman (Stephen Johnson

lab, WashU). For further information see  
[http://www.sanger.ac.uk/Projects/D\\_rerio/fishmask.shtml](http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml) DKEYP-86B5  
 is from a Zebrafish BAC library  
 VECTOR: pIndigoBAC-5.

FEATURES source  
 1. .223724  
 Location/Qualifiers  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="DKEYP-86B5"  
 /clone\_lib="DanioKeyypilot"

## ORIGIN

Query Match 75.2%; Score 18.8; DB 5; Length 223724;  
 Best Local Similarity 90.9%; Pred. No. 7.5e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTGCAA 25  
 ||||| ||||| ||||| ||||| |||||  
 Db 84895 ATTAAGAGAGACTATTTCGTGCAA 84874

## RESULT 43

## AC098905/c

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

AC098905 225849 bp DNA linear HTG 10-MAY-2003  
 Rattus norvegicus clone CH230-163J19, WORKING DRAFT SEQUENCE, 2  
 unordered pieces.

AC098905.6 GI:30521215

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridea; Muridae; Murinae; Rattus.

1 (bases 1 to 225849)

Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,

Anyaiebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,

Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,

Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Caldeiron, E.,

Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,

Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,

Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,

Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,

Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,

Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,

Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guavara, W.,

Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,

Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,

Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,

Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,

Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,

Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,

Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

Lorenzuehwa, L., Loulsegue, H., Lozano, R., Lu, X., Lu, X., Ma, J.,

Mareshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Mangum, B., Mapua, P., Martin, K., Martin, R., Martine, E.,

Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,

Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,

Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,

Nankarville, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,

Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,

Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,

Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.-L.,

Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,

Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,

Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,

Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sorengren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
JOURNAL

COMMENT

Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
Project Information  
Center project name: GIOC  
Center clone name: CH230-163J19  
Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 212145 bases at least Q40  
Consensus quality: 214496 bases at least Q30  
Consensus quality: 216260 bases at least Q20  
Estimated insert size: 223073; sum-of-contigs estimation  
Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved

1 222215: contig of 222215 bp in length  
222216 222315: gap of unknown length  
222316 225849: contig of 3534 bp in length.  
Location/Qualifiers  
1. .225849  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"

FEATURES  
source

/clone="CH230-163J19"  
1. .1302  
/note="wgs\_contig"  
214638. .217165  
/note="wgs\_contig"  
220987. .222215  
/note="wgs\_contig"  
222216. .222315  
/estimated\_length=unknown

misc\_feature

misc\_feature

misc\_feature

gap

ORIGIN

Query Match 75.2%; Score 18.8; DB 14; Length 225849;  
Best Local Similarity 90.9%; Pred. No. 7.5e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCTGTCG 23

Db 173065 AAATAAAGAGACTATTCTGTCG 173044

RESULT 44

BX682552/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [zfish-help@sanger.ac.uk](mailto:zfish-help@sanger.ac.uk)

Center: Wellcome Trust Sanger Institute

http://www.sanger.ac.uk/Projects/D\_rerio/fishmask.shtml DKEY-120C6  
is from a Zebrafish BAC library

VECTOR: pindigoBAC-5

Location/Qualifiers

1. 228044  
/organism="Danio rerio"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7955"  
/clone="DKEY-120C6"  
/clone\_lib="DanioKey"

#### FEATURES

source

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 5; Length 228044;

Best Local Similarity 90.9%; Pred. No. 7.4e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCTGTCAA 25

||||| |  
Db 44750 ATTAAGAAAGTATTCTGTCAA 44729

#### RESULT 45

AC120660/c

LOCUS

AC120660 231498 bp DNA linear HTG 20-NOV-2002

Rattus norvegicus clone CH230-33A2, WORKING DRAFT SEQUENCE, 2

unordered pieces.

AC120660

AC120660.6 GI:25137880

HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Murioidea; Muridae; Rattus;

1 (bases 1 to 231498)

Muzny,D,Marie., Metzker,M, Lee., Abramson,S., Amin,A., Anguiano,D.,

Allen,C., Allen,H., Alsbrooks,S., Ayodeji,M., Baca,E., Baden,H.,

Anyalebechi,V., Aoyagi,A., Ayodeji,M., Barnstead,M., Benahmed,P.,

Baldwin,D., Bandaranaike,D., Barber,K., Blyth,K., Calderon,E.,

Biswal,K., Blair,J., Blankenburg,K., Burch,P., Burrell,K., Center,A.,

Bryant,N., Buhay,C., Carter,K., Cavazos,I., Ceasar,H., Chen,R.,

Cardenas,V., Carter,K., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,

Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,

Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

Georgiev,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,

Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,

Harvey,Y., Havlak,P., Haves,A., Henderson,N., Hernandez,J.,

Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,

Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,

Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,

Karpthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,

Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,

Liu,J., Liu,Y., London,P., Longacre,S., Lopez,J.,

Lorenshewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,

Mareshwari,M., Mahindarte,M., Mahmoud,M., Malloy,K., Mangum,A.,

Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,

Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,

Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,

Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,

Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

Nwaokeleneh,O., Okunonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,

Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,

Plopper,F., Poldexter,A., Popovic,D., Primus,E., Pu,L.-L.,

Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,

Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,

Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,

Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,  
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,  
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,  
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,  
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,  
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,  
Williams,G., Willson,R., Wleciyk,R., Wooden,H., Worley,K.,  
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,  
Yu,F., Zhang,J., Zhou,X., Zhou,S., Zhao,S., Dunn,D., von  
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,  
Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 231498)

Worley,K.C.

Direct Submission

Submitted (09-MAY-2002)

Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 231498)

Rat Genome Sequencing Consortium.

Direct Submission

Submitted (20-NOV-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Nov 20, 2002 this sequence version replaced gi:23269608.

The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: http://www.hgsc.bcm.tmc.edu/

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GXLB

Center clone name: CH230-33A2

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 224823 bases at least Q40

Consensus quality: 225934 bases at least Q30

Consensus quality: 226530 bases at least Q20

Estimated insert size: 229373; sum-of-contigs estimation

Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).

\* NOTE: This is a 'working draft' sequence. It currently  
consists of 2 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

\* 1 229826: contig of 229826 bp in length

\* 229827 229926: gap of unknown length

\* 229927 231498: contig of 1572 bp in length.

----- Location/Qualifiers

1. 231498

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-33A2"

#### FEATURES

source

```

misc_feature 1. .1414
      /note="wgs_end_extension
clone_end:Sp6"
3437. .4234
misc_feature
      /note="clone_boundary
clone_end:Sp6
site:_end:Sp6"
end_sequence:BH351289"
complement(224954. .225701)
/note="clone_boundary
clone_end:T7
site:_end:T7"
end_sequence:BH351287"
227043. .229826
misc_feature
      /note="wgs_end_extension
clone_end:T7"
229827. .229926
      /estimated_length=unknown

ORIGIN
Query Match 75.2%; Score 18.8; DB 14; Length 231498;
Best Local Similarity 90.9%; Pred. No. 7.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AATTAAGAGACTATTCGTGC 23
|||||
Db 87242 AATTAAGAGACTATTCGTGC 87221

RESULT 46
BV515800/c
LOCUS
DEFINITION
      rpf75e02.g1 Clint Pan troglodytes versus STS genomic, sequence
      tagged site.
ACCESSION BV515800
VERSION BV515800
KEYWORDS BV515800.1 GI:623933558
SOURCE STS.
ORGANISM Pan troglodytes versus
      Pan troglodytes versus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
      Homnidae; Pan.
REFERENCE 1 (bases 1 to 575)
AUTHORS Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
      Jaffe,D.B.
TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the
JOURNAL Human Genome
COMMENT Unpublished (2005)
Contact: Michael C. Zody
Broad Institute of MIT and Harvard
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172580933
Fax: 6172580903
Email: mczody@broad.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 575
Protocol:
23,021,928 chimpanzee whole genome shotgun reads were aligned to
the Human genome NCBI
Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,
including Clint (Pan
troglodytes versus), 3 other Pan troglodytes versus chimps
(Donald, Karlien, Yvonne), 3 Pan
troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps
of unknown origin
(Gon, Unknown Chimp). Common names: Pan troglodytes versus is the
western chimp and Pan
troglodytes troglodytes is the central chimp. To be included in
chimpanzee SNP discovery, a
read must be at least 500bp in length, at least 50% of its base
calls must have Phred
score >= 20, at least 30% of its base calls must satisfy
SNQS(30,25)(single strand NQS, the
base in question has Phred score >= 30, the surrounding 10 bases in
the read have Phred
score >= 25), and the read must have at least 200 bp SNQS(30,25)
bases. Reads not uniquely
placed in the genome and read pairs whose two ends were not
consistently placed were
discarded. After above filtering, NQS(30,25) standard was applied
to all pairs of
overlapping reads to call NQS bases and SNPs. Alignments (between
two reads) with less
than 100 NQS bases or with SNP rate > 0.01 were discarded. To
exclude alignment between two
copies of a single read, comparisons between two reads that share
95% of their genome
alignments (>=95% bases of read A and >=95% bases of read B were
placed at the same locus
of human genome) were discarded.
FEATURES
      source
      Location/Qualifiers
      1..575
      /organism="Pan troglodytes versus"
      /mol_type="genomic DNA"
      /sub_species="verus"
      /db_xref="taxon:37012"
      /clone_lib="Clint"
      <1..575

STS
ORIGIN
Query Match 74.4%; Score 18.6; DB 10; Length 575;
Best Local Similarity 84.0%; Pred. No. 3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCAA 25
|||||
Db 419 CAAATTAAGAGACTATTCGTGCAA 395

RESULT 47
BV552932/c
LOCUS
DEFINITION
      S217P6940PB3.T0 Noemie Pan troglodytes troglodytes STS genomic,
      sequence tagged site.
ACCESSION BV552932
VERSION BV552932.1 GI:62443952
KEYWORDS STS.
SOURCE Pan troglodytes troglodytes
ORGANISM Pan troglodytes troglodytes
      Pan troglodytes troglodytes
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
      Homnidae; Pan.
REFERENCE 1 (bases 1 to 694)
AUTHORS Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and
      Jaffe,D.B.
TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the
JOURNAL Human Genome
COMMENT Unpublished (2005)
Contact: Michael C. Zody
Broad Institute of MIT and Harvard
320 Charles Street, Cambridge, MA 02141, USA
Tel: 6172580933
Fax: 6172580903
Email: mczody@broad.mit.edu
Primer A: No sequence submitted
Primer B: No sequence submitted
STS size: 694
Protocol:
23,021,928 chimpanzee whole genome shotgun reads were aligned to
the Human genome NCBI
Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,
including Clint (Pan
troglodytes versus), 3 other Pan troglodytes versus chimps
(Donald, Karlien, Yvonne), 3 Pan
troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps
of unknown origin
(Gon, Unknown Chimp). Common names: Pan troglodytes versus is the
western chimp and Pan
troglodytes troglodytes is the central chimp. To be included in
chimpanzee SNP discovery, a
read must be at least 500bp in length, at least 50% of its base
calls must have Phred

```

(Donald, Karlén, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score  $\geq 20$ , at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has Phred score  $\geq 30$ , the surrounding 10 bases in the read have Phred score  $\geq 25$ ), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate  $> 0.01$  were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments ( $> 95\%$  bases of read A and  $> 95\%$  bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

source  
Location/Qualifiers  
1..694  
/organism="Pan troglodytes troglodytes"  
/mol\_type="genomic DNA"  
/sub\_species="troglodytes"  
/db\_xref="taxon:37011"  
/clone\_lib="Noemie"  
<1..>694

STS  
ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 694;  
Best Local Similarity 84.0%; Pred. No. 2.9e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGCAA 25  
|||||  
Db 609 CAAATTAAAGAGATTATTAGCAA 585

RESULT 48  
BV529432/c  
LOCUS 703 bp DNA linear STS 08-APR-2005  
DEFINITION G591P613281P66.T0 Clint Pan troglodytes verus STS genomic, sequence tagged site.  
VERSION BV529432  
KEYWORDS BV529432.1 GI:62407216  
SOURCE STS.  
ORGANISM Pan troglodytes verus  
Pan troglodytes verus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.

REFERENCE 1 (bases 1 to 703)  
AUTHORS Mikkelsen, T.S., Hillier, W.L., Eichler, E.E., Zody, M.C. and Jaffe, D.B.  
TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome  
JOURNAL Unpublished (2005)  
COMMENT Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu

Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 703  
Protocol: 23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlén, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score  $\geq 20$ , at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has Phred score  $\geq 30$ , the surrounding 10 bases in the read have Phred score  $\geq 25$ ), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate  $> 0.01$  were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments ( $> 95\%$  bases of read A and  $> 95\%$  bases of read B were placed at the same locus of human genome) were discarded.

## FEATURES

source  
Location/Qualifiers  
1..703  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1..>703

STS  
ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 703;  
Best Local Similarity 84.0%; Pred. No. 2.9e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGCAA 25  
|||||  
Db 540 CAAATTAAAGAGATTATTAGCAA 516

RESULT 49  
BV521567/c  
LOCUS 940 bp DNA linear STS 08-APR-2005  
DEFINITION G591P66182RD10.T0 Clint Pan troglodytes verus STS genomic, sequence tagged site.

ACCESSION BV521567  
VERSION BV521567.1 GI:62399337  
KEYWORDS STS.  
SOURCE Pan troglodytes verus  
ORGANISM Pan troglodytes verus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan.

REFERENCE 1 (bases 1 to 940)  
AUTHORS Mikkelsen, T.S., Hillier, W.L., Eichler, E.E., Zody, M.C. and Jaffe, D.B.  
TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the



JOURNAL COMMENT  
Human Genome  
Unpublished (2005)

Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu  
Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 940  
Protocol:

23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald, Karlien, Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie, Masuku, Clara) and 2 chimps of unknown origin (Gon, Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score >= 20, at least 30% of its base calls must satisfy SNQS(30,25) (single strand NQS, the base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.

FEATURES  
source

1..940  
/organism="Pan troglodytes verus"  
/mol\_type="genomic DNA"  
/sub\_species="verus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1..940

STS  
ORIGIN

Query Match 74.4%; Score 18.6; DB 10; Length 940;  
Best Local Similarity 84.0%; Pred. No. 2.7e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCTGCA 25  
DB 665 CAATTAAGAGACTATTCTGCA 641

RESULT 50  
AY513464/c

LOCUS  
DEFINITION  
AY513464 1812 bp DNA linear INV 07-JUL-2004  
Stalotropa fairmairii isolate STAL cytochrome oxidase subunit I (COI) gene, partial cds; rRNA-Leu gene, complete sequence; and cytochrome oxidase subunit II (COII) gene, partial cds; mitochondrial.

ACCESSION

AY513464

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AV513464.1 GI:47119012  
mitochondrion Stalotropa fairmairii  
Stalotropa fairmairii

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Paraneoptera; Hemiptera; Euhemiptera; Membracoidea; Membracidae; Stalotropa.

REFERENCE  
AUTHORS  
TITLE

1 (bases 1 to 1812)  
Lin, C.-P., Danforth, B.N. and Wood, T.K.  
Molecular phylogenetics and evolution of maternal care in membracine treehoppers

JOURNAL  
AUTHORS  
TITLE

Syst. Biol. 53 (3), 400-421 (2004)  
2 (bases 1 to 1812)  
Lin, C.-P., Danforth, B.N. and Wood, T.K.  
Direct Submission

JOURNAL  
AUTHORS  
TITLE

Submitted (29-DEC-2003) Entomology, Cornell University, Comstock Hall, Ithaca, NY 14853-0901, USA

FEATURES  
source

1..1812  
Location/Qualifiers  
/organism="Stalotropa fairmairii"  
/organelle="mitochondrion"  
/mol\_type="genomic DNA"  
/isolate="STAL"  
/db\_xref="taxon:269982"  
<1..1232  
/gene="COI"  
<1..1232  
/gene="COI"  
/note="TAA stop codon is completed by the addition of 3' A residues to the mRNA"  
/codon\_start=2  
/transl\_except=(pos:1232,aa:TERM)  
/transl\_table=5  
/product="cytochrome oxidase subunit I"  
/protein\_id="AA111466.1"  
/db\_xref="GI:47119013"  
/translation="FLPESLILLTSTAGGAGTGTVPPLSSNVAHSGPSVDLVI  
PSLHAGASSILGAINLITIMNRCESKLDQSLFTWSIFITAFLLILSLPVLAGA  
ITMLGDFRNFNTSFFDSGGDPIYQHLFWFCHPEVYILILPFGVISHIVQESG  
KNSFGYIGMMYAMISIGGLGFIYWAHMFVTGMDVDTATYSTWIIALPTGIKIF  
SLWATMGSCLLKSLWSMIWAGSFVLFVGLTGVILVANSIDILHDITYYVAHFY  
VLSMGAVFTIMAGHIWYPLITGVTMNKMKIQFFSMFIGVNNMTPFPHFLGNGMP  
RRYSDYPSYMLNWNVSSFGSVLSISITFMIFVIESLMSKRLSLFSMNSSIEML  
QEYPPKHSYSELPLVN"

gene  
CDS

tRNA  
gene  
CDS

1233..1298  
/product="tRNA-Leu"  
1299..>1812  
/gene="COII"  
1299..>1812  
/gene="COII"  
/codon\_start=1  
/transl\_table=5  
/product="cytochrome oxidase subunit II"  
/protein\_id="AA111467.1"  
/db\_xref="GI:47119014"  
/translation="MANIKMSFODPSSNNMEOLVFENDHVMIMILITMVLVYLMSFS  
SMVSVLNRNLGQLIEFWLAPILLVFIAPILVPSLKILYLLETISPMITIKLGH  
QWYSYEFSDFTKIEFDSYMKPMNDLNQOEFKFLVDNRLVVPYIPIRILASSLDVV  
HSWTIPSLGVR"

ORIGIN

Query Match 74.4%; Score 18.6; DB 2; Length 1812;  
Best Local Similarity 84.0%; Pred. No. 2.4e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCTGCA 25

DB 1083 CTATTAAGAGACTACTTCCAA 1059

RESULT 51  
AY064171  
LOCUS

AY064171 2461 bp DNA linear BCT 26-DEC-2001



```

Streptococcus thermophilus pts operon, complete sequence.
DEFINITION
AY064111
VERSION
AY064111.1 GI:17978652
KEYWORDS
SOURCE
ORGANISM
Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 2461)
AUTHORS
van den Bogaard, P.T.C., van Wieringen, M., Memelink, K.,
Kleerebezem, M., Kuipers, O.P. and de Vos, W.M.
TITLE
Characterization of the Streptococcus thermophilus ptsHI operon and
analysis of HPr mutants
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 2461)
AUTHORS
van den Bogaard, P.T.C., van Wieringen, M., Memelink, K.,
Kleerebezem, M., Kuipers, O.P. and de Vos, W.M.
TITLE
Direct Submission
JOURNAL
Submitted (28-NOV-2001) Department of Flavour and Natural
Ingredients, Wageningen Centre for Food Sciences/NIZO Food
Research, P. O. Box 20, Ede, Gelderland 6710 BA, The Netherlands
Location/Qualifiers
FEATURES
1..2461
/organism="Streptococcus thermophilus"
/mol_type="genomic DNA"
/db_xref="taxon:1308"
309..2310
/gene="pts operon"
309..572
/gene="ptsH"
309..572
/feature="HPr"
309..572
/gene="ptsH"
/codon_start=1
/transl_table=1
/product="HPr"
/protein_id="AAL47557.1"
/db_xref="GI:17978653"
/translation="MASEKPHIVAEIIGHARPAITLLVQTSKFSASDITILEYKGAVAL
KSIQVMSIGVGGADVTIISABGPDAADIAIAETWTKEGLA"
577..2310
/gene="ptsI"
577..2310
/gene="ptsI"
/codon_start=1
/transl_table=1
/product="enzyme I"
/protein_id="AAL47558.1"
/db_xref="GI:17978654"
/translation="MTKMLKGIASDGVAVAKAYLLVQDLSFETVTVEDISAEEARL
DAALASQDELSVIREKAVESGEEAAAFDAHLMLVADPEMTGQIKETIRAKQVNAE
AALTEVTNMFIAIFEGMDNPNYQERAAADIVTKRVLANLLGKLPNPATIDEEI11
IHLVLPSTQALNKKFVKAFTVDIGRTSHSAIMARTLEIAAVLGINNITELVDGD
ILAVSGITGEVNTPEQITAEKPAAGEAYAKOAWAOLKDAPTVTADGKHFELAA
IGTPKVEGNDNGAEAGLVTPEFLYMSQDPTEEDQYEAQVLEGMGKPVVVR
ITMDIGDKGLPYDLKPNMPLCYLRALRISISETGNQFRTQLRALLRASVHGKLR
MFMVALLTEFRTKGLBEEKAVGAVVADDIETVGIMIEIPAAAMLADQFAKEV
DFFSITGNLQIYTWAAADRMEQVSYLYQPNPSILRLNNVIAKAAHAGKRWAGMCE
MAGDQTAVPLLVMGLMDFMSATSILRSLMKLDTAKMEYANRALTECSTMBEEV
LELSKKYVNDV"
ORIGIN
Query Match 74.4%; Score 18.6; DB 1; Length 2461;
Best Local Similarity 84.0%; Pred. No. 2.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 CAATTAAGAGACTATTCTGCGAA 25
| | | | | | | | | | | | | | | | | | | |
Db 841 CAATCAAGAAACTATCCGTCGTA 865
| | | | | | | | | | | | | | | | | | | |
RESULT 52
AY253327
Streptococcus thermophilus linear BCT 04-SEP-2003
(ptsH), and enzyme I (ptsI) genes, complete cds; and glyceraldehyde
3-phosphate dehydrogenase (gapN) gene, partial cds.
AY253327
AY253327.1 GI:30027104
Streptococcus thermophilus
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 3991)
AUTHORS
Cochu, A., Vadeboncoeur, C., Moineau, S. and Frenette, M.
TITLE
Genetic and Biochemical Characterization of the
Phosphoenolpyruvate:Glucose/Mannose Phosphotransferase System of
Streptococcus thermophilus
JOURNAL
Appl. Environ. Microbiol. 69 (9), 5423-5432 (2003)
REFERENCE
2 (bases 1 to 3991)
AUTHORS
Cochu, A., Vadeboncoeur, C., Moineau, S. and Frenette, M.
TITLE
Direct Submission
JOURNAL
Submitted (10-MAR-2003) Biochimie/Microbiologie, Universite Laval,
Cite Universitaire, Quebec, Que G1K 7P4, Canada
Location/Qualifiers
FEATURES
1..3991
/organism="Streptococcus thermophilus"
/mol_type="genomic DNA"
/strain="ATCC 19258"
/db_xref="ATCC:19258"
/db_xref="taxon:1308"
128..1303
/gene="icd"
128..1303
/gene="icd"
/codon_start=1
/transl_table=1
/product="isocitrate dehydrogenase"
/protein_id="AAP05988.1"
/db_xref="GI:30027105"
/translation="MAERILMTNGKLKVTNDPIIPFIEGGVGRDITWNAQAIKFKAV
EAYVQGRIMQEWELAGKAPDKTGEWLPQETBAIRESLVAIKGLPSTPVGGGIRS
LNVALQELDLVACVRYVYFEGVASPLKEPKETDITI PRENTEDIYAGIEWAGTVD
VERVIAFQTEMNVKIRFPKSSSIGIKPISIEGSKRLIRSAIDYAKNLNKKVTLVH
KGNIKFTGGPRKMGVEVAEEYEKEEMLAGLEVNDIADNPLQALLNKKPEVDVA
LITNLQNDYDALAAQVGIGISPGANINYQTHAIFEATHGTAPDIADODRANPCSI
LLSGCMLEIYIGWSEAAELIISAIEKNPKSGITFVDLAFKRAYSTSGFSNQILSI1"
1565..1828
/gene="ptsH"
1565..1828
/gene="ptsH"
/feature="phosphocARRIER"
/feature="phosphoenolpyruvate phosphotransferase system"
/codon_start=1
/transl_table=1
/product="HPr"
/protein_id="AAP05989.1"
/db_xref="GI:30027106"
/translation="MASEKPHIVAEIIGHARPAITLLVQTSKFSASDITILEYKGAVAL
KSIQVMSIGVGGADVTIISABGPDAADIAIAETWTKEGLA"
1833..3566
/gene="ptsI"
1833..3566
/gene="ptsI"
/feature="phosphotransferase system enzyme I;"
/feature="phosphoenolpyruvate phosphotransferase system (pts)"
/product="enzyme I"
/transl_table=1
/product="enzyme I"
/protein_id="AAP05990.1"
/db_xref="GI:30027107"
/translation="MTKMLKGIASDGVAVAKAYLLVQDLSFETVTVEDISAEEARL
DAALASQDELSVIREKAVESGEEAAAFDAHLMLVADPEMTGQIKETIRAKQVNAE
AALTEVTNMFIAIFEGMDNPNYQERAAADIVTKRVLANLLGKLPNPATIDEEI11

```

```

VAHDLTPSDTAQLNKKFKVAFVTDIGRTSHSAIMARTLEIPAVLGTNNITELVKDGD
ILVSGITGEVVIINPTESQIAIEFFKAAGEDYAKQKAEQAQLKADPTVTADGKHFELAN
IGTPKQGVNDNGAEAVGLYRTFLYMSQDPTTEQDYEAAYKAVLEGMNGKPVVVR
TMDIGGDKELPYFDLPKEMNPFGLGYRALRISISETGNQMFRTQRLALIRASVHGKJRI
MPPVALLTFRTRAKGILEEKAKILVAEGVAVANDIEVGINETIPAAALADQFAKEV
DPFSGTNDLLQYTMADRMNEQSYLYQYPNPSILRLINNVIRAAHAEGKAGWACGE
MAGDQAVPLLVGLDGFMSAISILTRSLMKKLDFAKMEEVANRALTECSMEEV
LELSKKYVNDV
3785..>3991
/gene="gapN"
3785..>3991
/gene="gapN"
/codon_start=1
/transl_table=11
/product="glyceraldehyde 3-phosphate dehydrogenase"
/protein_id="RAPO5991.1"
/db_xref="GI:30027108"
/translation="MAKOYKNVNGEWKTSSENSITIIYAPANGEELGSVPAMSAQAEVDE
VYAAKALPAWRALSYAERRAAYS"

ORIGIN
Query Match          74.4%; Score 18.6; DB 1; Length 3991;
Best Local Similarity 84.0%; Pred. No. 2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTGCTCAA 25
      ||||| ||||| ||||| |||||
Db 2097 CAAATCAAGAACTATCGTGCTA 2121

RESULT 53
BS0000014/c
LOCUS
DEFINITION
Pan troglodytes chromosome 22 clone:PTB-065D08, map 22, complete
sequences.
ACCESSION
BS0000014
VERSION
BS0000014.1 GI:37537281
KEYWORDS
HTG.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Pan.

REFERENCE
1
The International Chimpanzee Chromosome 22 Consortium.
DNA sequence and comparative analysis of chimpanzee chromosome 22
Nature 429, 382-388 (2004)
2
(bases 1 to 4243)
Hattori,M., Toyoda,A., Watanabe,H., Taylor,T.D., Kuroki,Y.,
Fujiyama,A. and Sakaki,Y.
Submitted (12-MAY-2003) Maashira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gscc.riken.go.jp, URL:http://hgsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
The Chimpanzee Chromosome 22 Sequencing Consortium consists of:
*Chinese National Human Genome Center at Shanghai, Shanghai, China;
*GBF, Dept. of Genome Analysis, Braunschweig, Germany; *Institute
of Molecular Biotechnology, Jena, Germany; *KRIBB Genome Research
Center, Daejeon, Korea;
*Max-Planck-Institute for Molecular Genetics, Berlin, Germany;
*National Institute of Genetics, Mishima, Japan;
*National Yang Ming University Genome Research Center, Taipei,
Taiwan;
*RIKEN Genomic Sciences Center, Yokohama, Japan.

----- Genomic Center
Center: RIKEN Genomic Sciences Center
Center code: RIKEN
Web site: http://hgsc.riken.go.jp/
Contact: hattori@gscc.riken.go.jp
----- Project Information
Center project name:The Chimpanzee Chromosome 22 Sequencing Project

```

```

Center clone name: PTB-065D08
----- Summary Statistics
Sequencing vector: pUC18,pUC13,prz19R; 100% of reads Chemistry:
Dye-terminator Big Dye and ET; 100% of reads Assembly program:
Phrap; version 0.990329
-----
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30);
an attempt was made to resolve all sequencing problems, such as
compressions and repeats; all regions were covered by at one
plasmid
subclone or more than one M13 subclone;
and the assembly was confirmed by restriction digest.
-----
Source information:
The RPCI-43 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/mchimp43.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac/).
VECTOR: pBACe3.6
The CHORI-251 chimpanzee BAC library was prepared from DNA isolated
from the blood of a single male chimpanzee using published
protocols (Osoegawa,K. et al. Genomics 52:1-8). The DNA from the
chimpanzee ('Clint') was obtained from the Yerkes Primate Center in
Atlanta. The library was prepared by Baoli Zhu, Chung Li Shu,
Kazutoyo Osoegawa, Evan Eichler & Pieter J de Jong. The library
characteristics are described at
http://www.chori.org/bacpac/chimpanzee251.htm.
The clone may be obtained from Pieter J. de Jong and coworkers
(http://www.chori.org/bacpac/).
VECTOR: pTARBAC2.1
The PTB1 chimpanzee BAC library was prepared from DNA isolated from
cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKS145
The PTB22 chimpanzee Fosmid library was prepared from DNA isolated
from cultured cells established from the blood of a single male
chimpanzee.
Clones may be obtained from Asao Fujiyama and co-workers
(http://www.gsc.riken.go.jp).
VECTOR: pKS143
-----
Sequence Quality Assessment:
This entry has been annotated with sequence
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in
10,000 bp.
-----
Neighboring clones: RP43-009002(left) and PTB-017F24(right).
Location/Qualifiers
1..4243
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="22"
/clone="PTB-065D08"
/clone_lib="PTB1 chimpanzee BAC"

FEATURES
source
Query Match          74.4%; Score 18.6; DB 8; Length 44243;
Best Local Similarity 84.0%; Pred. No. 1.3e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

ORIGIN

```

```

QY 1 CAAATTAAAGAGACTATTCTGTCAA 25
Db 24484 CAAATTAAAGAGACTATTCTGTCAA 24460

RESULT 54
WPCOMMENT
Sequence split into 5 fragments LOCUS CR735128 Accession CR735128
Fragment Name Begin End
CR735128_0 1 110000
CR735128_1 100001 210000
CR735128_2 200001 310000
CR735128_3 300001 410000
CR735128_4 400001 456610
Continuation (5 of 5) of CR735128 from base 400001 (CR735128 Danio rerio clone DK5Y-2010)

Query Match 74.4%; Score 18.6; DB 14; Length 56610;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTCAA 25
Db 29540 CAGATTAAAGAGACTATTCTGTCAA 29516

RESULT 55
AC017558/c
LOCUS AC017558 58651 bp DNA linear HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***.
ACCESSION AC017558
VERSION AC017558.1 GI:6554439
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 58651)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CD10211255 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..58651
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

ORIGIN
Query Match 74.4%; Score 18.6; DB 14; Length 58651;
Best Local Similarity 84.0%; Pred. No. 1.2e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTCAA 25
Db 55556 CAAATTACGGAACACTATTGTCGA 55532

RESULT 56
AC005421/c
LOCUS AC005421 69992 bp DNA linear INV 31-OCT-1998
DEFINITION Drosophila melanogaster, chromosome 2L, region 22A1-22A1, P1 clone
DS03601, complete sequence.
ACCESSION AC005421
VERSION AC005421.1 GI:3818344

```

```

KEYWORDS HTG.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 69992)
AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoakins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Masda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Poon, L., Poon, L., Poon, L., Poon, L.,
Svirskas, R.R., Twomey, B., Wan, K.H., Weinburg, T., Zhang, R.,
Zieran, L.L. and Rubin, G.M.
TITLE Sequencing of Drosophila chromosome 2L, region 22A1-22A1
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 69992)
AUTHORS Celniker, S.E., George, R.A., Galle, R., Svirskas, R.R., Hoskins, R.A.,
Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C.,
Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,
Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B.,
Lomotan, M.A., Mak, J., Masda, P., Mok, M.S., Moshrefi, A.R.,
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L.,
Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,
Zieran, L.L. and Kimmel, B.E.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Oct 31, 1998 this sequence version replaced gi:3805728.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to bdg@fruitfly.berkeley.edu.
P1 library location: 38-49
FEATURES
source
1..69992
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/strains="Y2: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="22A1-22A1"
/clone="P1 DS03601 (D334)"
/clone_lib="P1 library, partial Sau3A in pAdiosacB11"
/note="DS03601 (D334) is a partial bridge which extend
from P1 end at bp 1 and extends to bp 69,992, minimally
overlapping DS08613 (D65)."
ORIGIN
Query Match 74.4%; Score 18.6; DB 2; Length 69992;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTCAA 25
Db 68289 CAAATTACGGAACACTATTGTCGA 68265

RESULT 57
AC004439/c
LOCUS AC004439 84055 bp DNA linear INV 28-JUL-1998
DEFINITION Drosophila melanogaster DNA sequence (P1 DS08613 (D65)), complete
sequence.
ACCESSION AC004439 AC000837-AC000848 AC001344 AC001345 AC001346 AC001347
AC001836-AC001847 AC002137 AC002888 AC002889 AC002890 AC002891
L81558 L81559 L81560
VERSION AC004439.1 GI:3342722

```

KEYWORDS	HTG.	JOURNAL	and Salanoubat, M.
SOURCE	Drosophila melanogaster (fruit fly)	REFERENCE	Unpublished
ORGANISM	Drosophila melanogaster	TITLE	2 (bases 1 to 91303)
REFERENCE	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	AUTHORS	EU Arabidopsis sequencing project.
AUTHORS	1 (bases 1 to 84055) Celniker, S.E., George, R.A., Galle, R.F., Hoskins, R.A., Svirkas, R.R., Harris, N.L., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., PUNCH, E., Snir, E., Twomey, B., Wan, K.H., White, L., Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.E.	Submitted (04-FEB-2000)	MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG. E-mail: lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project Coordinator: Marcel Salanoubat and Francis Quetier, Grouperment d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr
TITLE	Sequencing of Drosophila chromosome 2L, region 22A1-22A3	COMMENT	Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <a href="http://www.mips.biochem.mpg.de/proj/thal/">http://www.mips.biochem.mpg.de/proj/thal/</a> .
JOURNAL	Unpublished (1997)	FEATURES	Location/Qualifiers
REFERENCE	2 (bases 1 to 84055)	1.	91303
AUTHORS	Celniker, S.E., George, R.A., Galle, R., Svirkas, R.R., Hoskins, R.A., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., PUNCH, E., Snir, E., Twomey, B., Wan, K.H., White, L., Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.E.	/organism="Arabidopsis thaliana"	
TITLE	Submitted (17-MAR-1998) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA	/mol_type="genomic DNA"	
JOURNAL	On Jul 28, 1998 this sequence version replaced gi:3334915.	/db_xref="taxon:3702"	
COMMENT	Sequence submitted by: Berkeley Drosophila Genome Project Lawrence Berkeley National Laboratory, MS 64-121 Berkeley, CA 94720	/chromosome="3"	
	For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site ( <a href="http://fruitfly.berkeley.edu/sequence/">http://fruitfly.berkeley.edu/sequence/</a> ) or send email to <a href="mailto:location@hgsc.lbl.gov">location@hgsc.lbl.gov</a> .	/ecotype="Columbia"	
FEATURES	Library location: 69-90.	1815..3883	
source	Location/Qualifiers	/gene="P2413.10"	
1..84055		join(1815..1871,2260..2448,2531..2839,2912..3145,3239..3883)	
/organism="Drosophila melanogaster"		/note="similarity to several hypothetical proteins - different species"	
/mol_type="genomic DNA"		/codon_start=1	
/db_xref="taxon:7227"		/product="putative protein"	
/chromosome="2L"		/protein_id="CAB72183.1"	
/map="22A1-22A3"		/db_xref="GI:6911863"	
/clone="P1 DS08613 (D65)"		/db_xref="GOA:Q9M1K5"	
		/db_xref="InterPro:IPR001594"	
		/db_xref="UniProt/TREMBL:Q9M1K5"	
		/translation="MANNEKLELYOVWRSNKFVCGRLIFGPDASSLVLTLLILGPVFFVKMTTKWADPTKPNLCIFILCVSWLTLTLDIFLLMTSRDPIVPRSPRPETDADPSTTSEMEVSGRTNIRIPRVKDVTVNGHTVKVFCDTCLLYRPRASHCISCNQVQFDFHCPWQVQICGVNRVFFMFISTTLCITVYFAPFSLNIFRPHMDKISIKWAKSKDVLSDILIVYCFITVWFVGLTIFHSYLCITNTTVENFRYDKKENYKIGILGNWEIIFLSKIPPSNMKFRSFVKEDYMMVMVETPTSLNGESLVSKKIDIEWGGRIVDESGKSYSLPEILNLNVEDLDDCEDDLKADHHTHHHHQHONHGIIPFPDFFFTNEIGSNKDERNGEESGSSDGTGKRVRSVDEDEKVEGYERNWS	
		1815..1871	
		/gene="P2413.10"	
		/number=1	
		1872..2259	
		/gene="P2413.10"	
		/number=1	
		2260..2448	
		/gene="P2413.10"	
		/number=2	
		2449..2530	
		/gene="P2413.10"	
		/number=2	
		2531..2839	
		/gene="P2413.10"	
		/number=3	
		2840..2911	
		/gene="P2413.10"	
		/number=3	
		2912..3145	
		/gene="P2413.10"	
		/number=4	
		3146..3238	
		/gene="P2413.10"	
		/number=4	
		3239..3883	
		/gene="P2413.10"	
		/number=5	
		4604..6279	
ORIGIN			
Query Match	74.4%; Score 18.6; DB 2; Length 84055;		
Best Local Similarity	84.0%; Pred. No. 1.1e+03;		
Matches	21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;		
QY	1 CAATTAAGACACTATTGTCGCAA 25		
Db	1068 CAATTAAGACACTATTGTCGCAA 1044		
RESULT 58			
ATF2413/C			
LOCUS	ATF2413 91303 bp DNA linear PLN 16-APR-2005		
DEFINITION	Arabidopsis thaliana DNA chromosome 3, BAC clone F2413.		
ACCESSION	AL138655		
VERSION	AL138655.1 GI:69999905		
KEYWORDS	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi		
REFERENCE	1		
AUTHORS	Nyakatura, G., Fartmann, B., Dauner, D., Sterr, W., Holland, R., Weichselgartner, M., Mewes, H.W., Lemcke, K., Mayer, K.F.X., Quetier, F.		

CDS  
/gene="F2413.20"  
Join(4604. .5077,5292. .5356,5433. .5760,5838. .5936,  
6016. .6279)  
/gene="F2413.20"  
/note="in the published mRNA sequence a downstream lying  
ATG at position 4712 is used as start codon"  
/codon\_start=1  
/product="leucine zipper-containing protein AT103"  
/protein\_id="CAB72164.1"  
/db\_xref="GI:6911864"  
/db\_xref="GOA:Q9MIK4"  
/db\_xref="InterPro:IPR008434"  
/db\_xref="UniProt/TREMBL:Q9MIK4"  
/translation="MAAENALVKPISKFSFSPKLSNPSPKFLSGRRFSTVIRMSASSPP  
PTTATSKSGTKKEIOESLTPRFTYDDFEMLQFNTEINKNLNEAEFALLQKEF  
KTDVQTHFVRNKEKRAADKLGPLROI FVFLERSCTABFSGFLLYKELGRRLLQKT  
NPVVAIEFSLSRDEARHAGFLNKLSDFNALDLGLFTKARKYTFEPKPFVATYL  
SEKIGYRITTIYRHLKENPEFCYPIFKYFENWCODENRHGDFFSALMAQOPFLND  
WQAKLWSRFCLSVVTYMYLNDQRTNFEYEGILNTEKFDMEHVIETNRTTAIFPAV  
LDVENPEFKRLDRMVVSYEKLLAIGETDDASFIKTLKRIPLVTSLSASEILAAVLMPP  
VSSGSDRAEEFEPNLVY"  
4604. .5077  
/gene="F2413.20"  
/number=1  
5078. .5291  
/gene="F2413.20"  
/number=1  
5292. .5356  
/gene="F2413.20"  
/number=2  
5357. .5432  
/gene="F2413.20"  
/number=2  
5433. .5760  
/gene="F2413.20"  
/number=3  
5761. .5837  
/gene="F2413.20"  
/number=3  
5838. .5936  
/gene="F2413.20"  
/number=4  
5937. .6015  
/gene="F2413.20"  
/number=4  
6016. .6279  
/gene="F2413.20"  
/number=5  
6596. .7654  
/gene="F2413.30"  
complement(Join(6596. .6760,6849. .7109,7367. .7654))  
/gene="F2413.30"  
complement(Join(6596. .6760,6849. .7109,7367. .7654))  
/gene="F2413.30"  
/note="similarity to NODULIN-26 - Glycine max,  
ENBL:X04782"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB72165.1"  
/db\_xref="GI:6911865"  
/db\_xref="GOA:Q9MIK3"  
/db\_xref="InterPro:IPR000425"  
/db\_xref="UniProt/Swiss-Prot:Q9MIK3"  
/translation="MGRIGLVVDLVLSFNIWAGVLVNLVHGVLFGRFTDPSGRIV  
RYLFHSIIVPEIFAYLAQATKGLNPLTALAAGVSGGFSFIPSVFVRIPVEVIGSIL  
AVKHIIHVPEIKGKPLNVAIHGALTGILFTFVLLSMGLTRKIPGSPFMTWIG  
SLAKTLHIIGSLTGCXMPAAVMGWAYARGEHITKEHLLVYVWLGPKVATLLAVWVF  
KVFKPLTERQEPKAKSE"  
complement(6596. .6760)  
/gene="F2413.30"  
/number=1  
complement(6761. .6848)  
/gene="F2413.30"

exon  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon  
intron  
exon  
gene  
gene  
CDS  
exon  
intron

/number=1  
complement(6849. .7109)  
/gene="F2413.30"  
/number=2  
complement(7110. .7366)  
/gene="F2413.30"  
/number=2  
complement(7367. .7654)  
/gene="F2413.30"  
/number=3  
8967. .11895  
/gene="F2413.40"  
Join(8967. .10168,10257. .10424,10510. .10675,10756. .10872,  
10961. .11154,11233. .11363,11450. .11522,11607. .11895)  
/gene="F2413.40"  
/note="strong similarity to  
phosphatidylinositol-4-phosphate 5-kinase AtPIP5K1 -  
Arabidopsis thaliana, EMBL:AB005902  
Contains ATP/GTP-binding site motif A (P-loop) AA102-109"  
/codon\_start=1  
/product="phosphatidylinositol-4-phosphate 5-kinase-like  
protein"  
/protein\_id="CAB72166.1"  
/db\_xref="GI:6911866"  
/db\_xref="GOA:Q9MIK2"  
/db\_xref="InterPro:IPR002498"  
/db\_xref="InterPro:IPR003409"  
/db\_xref="UniProt/TREMBL:Q9MIK2"  
/translation="MSKQSCVLKAWETVRKTQQAKKRANSIFGTVSVAPODDDAT  
TYTERDDTSTNRSSIGELIYHAERILPNGDYVTGOWYDSFPHGHGKYLWTGCMYIG  
DWYNGKTWGRGKFWPSPGATYEGEFKSGMDGVGLYTGPSGDTYKQWYMNLLKHGHI  
KRFANGDYVGEWRGLQEAQGYQWRDGSYYTGMWKNATI CGKGTFFIWDGNRYDGF  
WDDGPRNGTGFKWADGSFYVGHWSNDPEEMNGTYPPSGDDGSPENDPKDVFNTLSEY  
KICSGERVPLPSOKLSVWNSKVEKPRRMSVDGRVSGVDRAPEKNNMGTESE  
GAADIDSTRDLDAEMMLEAEFGIQLRPSAPMRLPRAGKQGETTSKGRNYVEL  
MNLQIGIRHAGVKQAPVVSLLDKHSADPPKEKVTREFEKTCTPPHQSSFFKWKD  
YCPVFRSLRKLKFKVDPADYMLSLICNDALRSLSPGKSGSFYLTNDRMYLTKMKK  
SETKVLRLMLAAYNNHRAFENTLIRFYGLHCVKLTGPLOKKRVFVWNLPCSEYS  
IHRFDLKGSSLGRTTDPSEINSNTILKLDLNFIFRLQKAWQEFIRQVMDKCEY  
LEQERIMDYSLAVGHFREASVAGELIPSGARTPIGEFESAPRLSRADVDQLLSDP  
TFWASIRLGGNNPARAERTMRRSDCEFLVGEPTGEYEVVWVIFGIIDILQDYDISKK

Query Match 74.4%; Score 18.6; DB 15; Length 91303;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CAANTTAAGACACTATTCTGCAA 25  
Db 81795 CAAAAAACAGAGACTATTGTGCAA 81771

RESULT 59  
AC002407/c  
LOCUS  
DEFINITION Human Chromosome X, complete sequence.  
AC002407  
AC002407.1 GI:2323254  
VERSION HTG.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
1 (bases 1 to 95506)  
Chen, E., Brownstein, B. H., States, D. J., Schlessinger, D. and  
Mazzarella, R.  
Direct Submission  
Unpublished (1997)  
2 (bases 1 to 95506)  
Brownstein, B. H., States, D. J. and Mazzarella, R.  
Direct Submission  
Submitted (12-AUG-1997) Center for Genetics in Medicine, Box 8232,  
Washington University School of Medicine, 4566 Scott Avenue, St.

COMMENT  
Louis, MO 63110, USA  
Current status of this project is available at:  
'http://genome.wustl.edu/cgm/seq\_projects.html',  
Submitted by:  
Elliason Chen,  
Advanced Center for Genetic Technology,  
Applied Biosystems Division of Perlin Elmer Corp.,  
850 Lincoln Center Drive,  
Foster City, CA 94404 USA  
e-mail: ellison@genseq.apldbio.com

and  
Buddy Brownstein,  
Center for Genetics in Medicine,  
Washington University School of Medicine, Box 8232  
4566 Scott Avenue,  
St. Louis, MO 63110, USA  
e-mail: buddy@genetics.wustl.edu

and  
David J. States,  
Institute for Biomedical Computing  
Washington University in St. Louis  
700 South Euclid Ave.  
St. Louis, MO 63108 USA  
e-mail: states@ibc.wustl.edu.

FEATURES  
source  
1..95506  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="bWDX177"

ORIGIN  
Query Match 74.4%; Score 18.6; DB 8; Length 95506;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGACACTATTTCGTGCA 25  
Db 85341 CAATTAAGACACTATTTCGTGCA 85317

RESULT 60  
AL451127 108958 bp DNA linear PRI 18-MAY-2005  
LOCUS Human DNA sequence from clone RP11-12817 on chromosome 9 Contains a  
DEFINITION novel gene, complete sequence.  
ACCESSION AL451127 AC025854  
VERSION AL451127.6 GI:14018295  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 108958)  
Brown, J.

REFERENCE  
AUTHORS Direct Submission  
TITLE Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
Clone requests: clonesrequest@sanger.ac.uk  
On or before May 15, 2001 this sequence version replaced  
gi:8077006, gi:13751503.

COMMENT  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human

chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr9  
RP11-12817 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBAC3.6  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----

Draft Sequence Produced by Whitehead Institute/MIT Center for  
Genome Research, 320 Charles Street,  
Cambridge, MA 02141, USA  
http://www-seq.wi.mit.edu  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.

FEATURES  
Location/Qualifiers  
source  
1..108958  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="9"  
/clone="RP11-12817"  
/clone\_lib="RPCI-11.1"  
100  
misc\_feature  
/note="Clone right end: RP11-404E6"  
join(92855..93084,97109..97192,98429..98738)  
/locus\_tag="RP11-12817.1-001"  
join(92855..93084,97109..97192,98429..98738)  
/locus\_tag="RP11-12817.1-001"  
/locus\_tag="RP11-12817.1-001"  
/note="match: ESTs: BM718494.1"  
108958  
misc\_feature  
/note="Clone right end: RP11-12817"

ORIGIN  
Query Match 74.4%; Score 18.6; DB 8; Length 108958;  
Best Local Similarity 84.0%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGACACTATTTCGTGCA 25  
Db 76723 CAATTAAGACACTATTTCGTGTA 76747

Search completed: January 27, 2006, 23:14:00  
Job time : 997.18 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:46:18 ; Search time 784.944 Seconds  
(without alignments)  
1448.344 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaaatgcagaacagg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

GenBank:

1: gb\_ba.\*  
2: gb\_in.\*  
3: gb\_env.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pr.\*  
9: gb\_ro.\*  
10: gb\_sts.\*  
11: gb\_sy.\*  
12: gb\_un.\*  
13: gb\_vi.\*  
14: gb\_htg.\*  
15: gb\_pl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	6	CS109173 Sequence
2	20	100.0	1731	6	CQ647508 Sequence
3	20	100.0	1734	6	AX608629 Sequence
4	20	100.0	21781	1	AE014229 Streptoco
5	20	100.0	29072	6	AX602188 Sequence
6	20	100.0	167050	1	SAG766847 Streptoco
7	20	100.0	349980	6	CQ655069 Sequence
8	20	100.0	349980	6	AX954529 Sequence
9	18.4	92.0	396	13	HIVAJ8558 Human imm
10	18.4	92.0	396	13	HIVAJ8564 Human imm
11	18.4	92.0	396	13	HIVAJ8568 Human imm
12	18.4	92.0	396	13	HIVAJ8582 Human imm
13	18.4	92.0	396	13	HIVAJ8589 Human imm
14	18.4	92.0	396	13	HIVAJ8639 Human imm
15	18.4	92.0	396	13	HIVAJ8642 Human imm
16	18.4	92.0	396	13	HIVAJ8646 Human imm
17	18.4	92.0	396	13	HIVAJ8656 Human imm
18	18.4	92.0	398	13	HIVAJ8543 Human imm

19	18.4	92.0	445	13	AF395561	AF395561 HIV-2 iso
20	18.4	92.0	708	13	HIVFAP	L33089 Human immun
21	18.4	92.0	9472	13	HIV2D194	J04542 Human immun
22	18.4	92.0	9472	13	HIVD194	X52223 Human immun
23	18.4	92.0	9473	6	A09995	A09995 HIV-2 (HIV-
24	18.4	92.0	10312	13	HIV2U38293	U38293 Human immun
25	18.4	92.0	10359	13	HIV2BEN	M30502 Human immun
26	18	90.0	37676	8	AY341245	AY341245 Homo sapi
27	18	90.0	110000	15	CR382125_04	Continuation (5 of
28	18	90.0	114519	8	AL157813	AL157813 Human DNA
29	18	90.0	194045	9	AL833774	AL833774 Mouse DNA
30	18	90.0	204712	14	AC134627	AC134627 Rattus no
31	18	90.0	276881	14	AC095481	AC095481 Rattus no
32	17.4	87.0	491	10	G52355	G52355 SHOC-85070
33	17.4	87.0	549	10	G90172	G90172 S210P60318B
34	17.4	87.0	723	10	BV558070	BV558070 qmm55f08.
35	17.4	87.0	737	10	BV586373	BV586373 G591P6318
36	17.4	87.0	800	10	BV576492	BV576492 G591P6083
37	17.4	87.0	800	10	BV628207	BV628207 S217P6230
38	17.4	87.0	806	10	BV594335	BV594335 taw08f01.
39	17.4	87.0	2081	5	BC063923	BC063923 Xenopus t
40	17.4	87.0	2286	8	AK123724	AK123724 Homo sapi
41	17.4	87.0	2303	2	AF401640	AF401640 Naegleria
42	17.4	87.0	2312	2	AY075353	AY075353 Drosophill
43	17.4	87.0	2345	2	AY089451	AY089451 Drosophill
44	17.4	87.0	4310	2	AF051933	AF051933 Drosophill
45	17.4	87.0	4911	6	CQ585744	CQ585744 Sequence
46	17.4	87.0	7886	6	CQ585743	CQ585743 Sequence
47	17.4	87.0	27103	14	AL353656	AL353656 Homo sapi
48	17.4	87.0	37170	8	HSB12389	Z68224 Human DNA 8
49	17.4	87.0	39601	15	AC079678	AC079678 Arabidops
50	17.4	87.0	50821	15	AC020579	AC020579 Arabidops
51	17.4	87.0	51405	14	AC019512	AC019512 Drosophill
52	17.4	87.0	54288	14	AC100240	AC100240 Mus muscu
53	17.4	87.0	63540	14	AC101468	AC101468 Mus muscu
54	17.4	87.0	65756	8	AL590286	AL590286 Human DNA
55	17.4	87.0	69817	15	AC003096	AC003096 Arabidops
56	17.4	87.0	84987	8	AL662897	AL662897 Human DNA
57	17.4	87.0	101526	14	AC106751	AC106751 Homo sapi
58	17.4	87.0	104632	8	AL136124	AL136124 Human DNA
59	17.4	87.0	110000	1	CR543861	Continuation (33 o
60	17.4	87.0	110000	14	CR848719_2	Continuation (3 of

## ALIGNMENTS

RESULT 1	CS109173	Sequence 4 from Patent EP1541697.	20 bp	DNA	linear	PAT 22-JUN-2005
LOCUS	CS109173	Sequence 4 from Patent EP1541697.				
DEFINITION	CS109173	synthetic construct				
ACCESSION	CS109173	other sequences; artificial sequences.				
VERSION	CS109173.1	GI:68148022				
KEYWORDS		synthetic construct				
SOURCE		other sequences; artificial sequences.				
ORGANISM		1				
REFERENCE		Unl,J.R., Cockerill,F.R., Aichinger,C. and Reiser,A.				
AUTHORS		Detection of group B streptococcus				
TITLE		Patent: EP 1541697-A 4 15-JUN-2005;				
JOURNAL		Mayo Foundation for Medical Education and Research (US)				
FEATURES		Location/Qualifiers				
source		1..20				
		/organism="synthetic construct"				
		/mol_type="unassigned DNA"				
		/db_xref="taxon:32630"				
		/note="Oligonucleotide"				

## ORIGIN

Query Match 100.0%; Score 20; DB 6; Length 20;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 1 CAAGTAAATGCAGAAACAGG 20

RESULT 2
CQ647508 1731 bp DNA linear PAT 02-FEB-2004
LOCUS Sequence 4465 from Patent WO0234771.
DEFINITION
ACCESSION CQ647508
VERSION CQ647508
KEYWORDS CQ647508.1 GI:41683480
SOURCE
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Telford, J., Masignani, V., Margarit y Ros, I., Grandi, G., Fraser, C.
and Tettelin, H.
TITLE Nucleic acids and proteins from streptococcus groups a & b
JOURNAL Patent: WO 0234771-A 4465 02-MAY-2002;
Chiron S.p.A. (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1..1731
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 1731;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 292 CAAGTAAATGCAGAAACAGG 311

RESULT 3
AX608629 1734 bp DNA linear PAT 17-FEB-2003
LOCUS Sequence 6558 from Patent WO02092818.
DEFINITION
ACCESSION AX608629
VERSION AX608629.1 GI:28404206
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Glaser, P., Rueniok, C., Chevalier, F., Frangeul, L., Lalioui, L.,
Zouine, M., Couve, B., Buchrieser, C., Poyart, C., Trieu-Cuot, P. and
Kunst, F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 6558 21-NOV-2002;
INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source
1..1734
/organism="Streptococcus agalactiae"
/mol_type="unassigned DNA"
/db_xref="taxon:1311"

ORIGIN
Query Match 100.0%; Score 20; DB 6; Length 1734;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 292 CAAGTAAATGCAGAAACAGG 311

RESULT 4
AE014229 21781 bp DNA linear BCT 23-SEP-2002
LOCUS Streptococcus agalactiae 2603V/R section 39 of 100 of the complete
DEFINITION genome.
ACCESSION AE014229 AE009948
VERSION AE014229.1 GI:22533831
KEYWORDS
SOURCE
ORGANISM Streptococcus agalactiae 2603V/R
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
1 (bases 1 to 21781)
Tettelin, H., Masignani, V., Cieslewicz, M. J., Eisen, J. A.,
Peterson, S., Wessels, M. R., Paulsen, I. T., Nelson, K. E., Margarit, I.,
Read, T. D., Madoff, L. C., Wolf, A. M., Beanan, M. J., Brinkac, L. M.,
Daugherty, S. C., DeBoy, R. T., Durkin, S., Kolonay, J. F., Umayam, L. A.,
Madupu, R., Lewis, M. R., Radune, D., Fedorova, N. B., Scanlan, D.,
Khouri, H., Mulligan, S., Cart, H. A., Cline, R. T., Gill, J.,
Scarselli, M., Mora, M., Iacobini, E. T., Brettoni, C., Galli, G.,
Mariani, M., Vegni, F., Maione, D., Rinaudo, D., Rappuoli, R.,
Telford, J. L., Kasper, D. L., Grandi, G. and Fraser, C. M.
Complete genome sequence and comparative genomic analysis of an
emerging human pathogen, serotype V Streptococcus agalactiae
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
12200547
2 (bases 1 to 21781)
Tettelin, H., Masignani, V., Cieslewicz, M. J., Eisen, J. A.,
Peterson, S., Wessels, M. R., Paulsen, I. T., Nelson, K. E., Margarit, I.,
Read, T. D., Madoff, L. C., Wolf, A. M., Beanan, M. J., Brinkac, L. M.,
Daugherty, S. C., DeBoy, R. T., Durkin, S., Kolonay, J. F., Umayam, L. A.,
Madupu, R., Lewis, M. R., Radune, D., Fedorova, N. B., Scanlan, D.,
Khouri, H., Mulligan, S., Cart, H. A., Cline, R. T., Gill, J.,
Scarselli, M., Mora, M., Iacobini, E. T., Brettoni, C., Galli, G.,
Mariani, M., Vegni, F., Maione, D., Rinaudo, D., Rappuoli, R.,
Telford, J. L., Kasper, D. L., Grandi, G. and Fraser, C. M.
Direct Submission
Submitted (18-JUL-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..21781
/organism="Streptococcus agalactiae 2603V/R"
/mol_type="genomic DNA"
/strain="2603V/R"
/serotype="V"
/db_xref="taxon:208435"
103..348
/gene="SAG0813"
103..348
/gene="SAG0813"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM99700.1"
/db_xref="GI:22533832"
/translation="MNNQINHCFSIMVLGVGTQMKKEFITGRMVARVMTLVCLCMV
SNLVLDMDIITEINEIVFCVLVASIAGLEFLQKERE"
387..674
/gene="SAG0814"
387..674
/gene="SAG0814"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM99701.1"
/db_xref="GI:22533833"
/translation="MTYIGVPKEQKALSRRGKQLEKINQGGKTLMSQIMQIDLVGSTCM
IYLEDGSOTHPYEILKVDVDWVQIRNRNKRSEEVIRIVRIHDIKELRPVK"
719..934
gene

```



```
/gene="SAG0815"
/719..934
/gene="SAG0815"
/notes="identified by match to PFAM protein family HMM
pf00381"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Cro/Ci family"
/protein_id="AAM99702.1"
/db_xref="GI:22533834"
/translation="WGLILKNRLKELARDGINQTELAKLTCVSRQTSLIERGEVTP
SVIAWKISQVFKESLENVFLSVSME"
934..1695
/gene="SAG0816"
934..1695
/gene="SAG0816"
/notes="similar to OMN1:NTL01SPL1088, identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAM99703.1"
/db_xref="GI:22533835"
/translation="MMKNFIKQSDTTFVRIIISLLIGGFIGAILSGVGFIIFGQDK
YISEINIVQFLWVRISVIVITAIPLSFILYIQIKYQKRVFNVDSESEIYRQINLR
HSYGMTFVIGISIVLNTLVNFKLNFDDSVTLVPIYDLSLLFVLGLHIIYFLKVI
RNIRGFKMTVAPTLEKLNKLQDAELESNYKMFCDIWMNLSGFIPTIYFVLFFI
SPVFOKVEIVAIITTSIHIIYILIKLSKAARHFR"
complement(1780..2343)
/gene="SAG0817"
complement(1780..2343)
/gene="SAG0817"
/notes="similar to OMN1:NTL01SPL0446, identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, putative"
/protein_id="AAM99704.1"
/db_xref="GI:22533836"
/translation="MSKNNNTCLIIETAIFAALAMALSMIPDASWFTPSFGAIPIL
FALRRRTKGLFAGLWGLHVLKSLVLSQVFEIYLAFISNGLAGVPSAKFKD
ALSSSKTKALSALSGAILATLVRYVWHYIAGVIFWASYPKMGSLTYSUSVNGTA
GLILFLFVVISIILVISYPSFLPKK"
complement(2584..3543)
/gene="nrdf-2"
/notes="synonym: SAG0818"
complement(2584..3543)
/gene="nrdf-2"
/notes="identified by match to PFAM protein family HMM
pf00268"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, beta
subunit"
/protein_id="AAM99705.1"
/db_xref="GI:22533837"
/translation="MTYTYEAINWNIEDVIDKSTWEKLTQFWLDTRIPLSLDLDDM
RKLSAQEKDLGVGFGLTLLDTMQSETGVGEIARADVRTPHEEAVLNNIQFMSVRAK
SYSIFSTLTKSEIEEIEFWTNNEFLQEKARIINDIYANGNALQKVKASTYLETFL
FYSGFTPLYLGNLKNLANVAEIIKLIIRDSEVHGTYIGYFQLGFNELPEDQENFR
DMYDLYLYENEKTKTYLDGVGWTVEVMTFLRYNANKALMNLQDPLFPDPTAND
VNFIVMNGISTGTSHNDFFSQVGNGLGSVEAHDDDDNYNGL"
complement(3746..5905)
/gene="nrde-2"
/notes="synonym: SAG0819"
complement(3746..5905)
/gene="nrde-2"
/notes="identified by match to PFAM protein family HMM
pf00317"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2, alpha
subunit"
```

```
/protein_id="AAM99706.1"
/db_xref="GI:22533838"
/translation="MSLKNIGDVSYFLRNLNEINRPVNGQIPLHKDKALTAFFKENVQ
PNSKAFDSITDKIAYLLKYDYLEEAFLNKRYPEFIEELSTKLFDKKFRKSPMAAYKF
QOYALKTNDGEYLYLESIEDRVLFNALYFADGDEELATDLALEMISQRPATPPGFLN
AGSRGELVSCFLIQVTDMMNAIGRSINSAQLSRIGGGVGISLSNLRACAGPIKGF
AGAAGVVPVVKLFEDSPSYSNQLGQROGAGVYLDVDFHPDIIISLSTKKNADKVR
VATLSGITVPDKFYELARNQEMTLFSPYSIEREYGVFFSIDITEKDELVANPNI
TKTKINARDLETEISKLQOESGYPYIINIDTANRPVVDGKLIMNLCSEILQVQKPS
LINDAQEYLEMTDISCNLSTNVLMNMTSPDGGKIKITMTRALFTVDDSSNIEAVPT
IKNGQAHTFGLGAMGLHSLAKNHIYGPSPESIEFTDIYFMLNMYTWLVSNNIAR
ERQTFVQFEKSYADGTYFDKYVSGKFVQSDVKVSLFANHFIPEAKDWNLRVAM
KDGLYQNR LAVAPNGSISYINDCSASIHPIQRIEERQEKIKIGIYYPANGLATDTI
PYTTSAYDMRMKRVLDVAAATEHVDQGLSMTLFLRSELKELYEWKTESKQTTTRDLS
ILRNAYAFNMKVKSIIYIRFTFTDDGSEVGANQCESVI"
complement(5983..6207)
/gene="nrdrh"
/notes="synonym: SAG0820"
complement(5983..6207)
/gene="nrdrh"
/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase 2,
nrdrh-redoxin"
/protein_id="AAM99707.1"
/db_xref="GI:22533839"
/translation="MAAITVFSKNNCMQCKMTKKFLDQHGADPFEINIDEKPEKIEYV
KNLGSAAAPVIEAGNVVFSGFQPSKLKELV"
6590..6853
/gene="ptsh"
/notes="synonym: SAG0821"
6590..6853
/gene="ptsh"
/notes="identified by match to PFAM protein family HMM
pf00381"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AAM99708.1"
/db_xref="GI:22533840"
/translation="WASKDPFHVAETGIIHARPATLAVQTASKFASDITLDYKGAENVL
KSIQVMSLGVCGQADVTISAGGADDDAIAAIEETWTKGLA"
6858..8591
/gene="ptsl"
/notes="synonym: SAG0822"
6858..8591
/gene="ptsl"
/notes="identified by match to PFAM protein family HMM
pf00391"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAM99709.1"
/db_xref="GI:22533841"
/translation="MTEMLKGIASAASGVAVAKAYLVLQPDLSPTETVVDNABEARL
DVALAQSDQLSVIREKAVESGAAAVFADHMLVLSDPEDMINQIKETIRAKQVNAE
TGLKEVTDMFITIFEGMEDNPYMQEAAIDRDVAKRVLALHGLVGLPNPATINEESIV
IAHDLTPSTATLAKQFVKAFVFNIGRSTSHSAIMARTLEIAAVALGTNDITERVQDQ
```

```
Query Match 100.0%; Score 20; DB 1; Length 21781;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAACAGG 20
|||
Db 7149 CAAGTAATGCAGAACAGG 7168

RESULT 5
AX602188 29072 bp DNA linear PAT 17-FEB-2003
LOCUS AX602188
DEFINITION Sequence 117 from Patent WO02092818.
ACCESSION AX602188
```







/translation="DHLVSGIRQVLFLEKIEPAQBEHEKYHNNKELSHKFGFLPTLV  
ARQVNTCAQCKQKGAHQVNAEIGTWQMDCTHLEKGIIVAVHVASGFEAEVVP  
QESGRQTALFLKLASRWPTIHLTDNGA"

## ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 396;  
Best Local Similarity 95.0%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20

Db 194 CAAGTAAATGCAGAAATAGG 213

## RESULT 11

HIVAJ8568  
LOCUS HIVAJ8568 396 bp DNA linear VRL 15-APR-2005  
DEFINITION Human immunodeficiency virus type 2 pol gene, isolate b1053.  
ACCESSION AJ008568  
VERSION AJ008568.1 GI:3286892  
KEYWORDS DNA polymerase; pol gene.  
SOURCE Human immunodeficiency virus 2 (HIV-2)  
ORGANISM Human immunodeficiency virus 2

Viruses; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.

## REFERENCE

1 Xiang,Z., Ariyoshi,K., Wilkins,A., Dias,F., Whittle,H. and

## AUTHORS

Breuer,J.  
HIV type 2 pathogenicity is not related to subtype in rural Guinea

## TITLE

Bisseau  
AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)

## JOURNAL

PUBMED  
9100992

## REFERENCE

2 Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F.,

## AUTHORS

Whittle,H. and Breuer,J.  
Correlation of HIV-2 genotype with progression to AIDS in vivo

## TITLE

Unpublished  
3 (bases 1 to 396)

## REFERENCE

Xiang,Z.  
Direct Submission

## AUTHORS

Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,  
Department of Virology, Queen Mary Westfield, Turner Street,

## TITLE

London, E1 2AD, UK

## JOURNAL

FEATURES

Location/Qualifiers

1..396

/organism="Human immunodeficiency virus 2"

/mol\_type="genomic DNA"

/strain="A"

/isolate="b1053"

/db\_xref="taxon:11709"

<1..>396

/gene="pol"

<1..>396

/gene="pol"

/codon\_start=2

/product="DNA polymerase"

/protein\_id="CAA08032.1"

/db\_xref="GI:3286893"

/db\_xref="GOA:Q9YQL5"

/db\_xref="InterPro:IPR001584"

/db\_xref="UniProt/TREMBL:Q9YQL5"

/translation="DHLVSGIRQVLFLEKIEPAQBEHEKYHNNKELSHKFGFLPTLV  
ARQVNTCAQCKQKGAHQVNAEIGTWQMDCTHLEKGIIVAVHVASGFEAEVVP  
QESGRQTALFLKLASRWPTIHLTDNGA"

ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 396;  
Best Local Similarity 95.0%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20

Db 194 CAAGTAAATGCAGAAATAGG 213

## ORIGIN

HIVAJ8568  
LOCUS HIVAJ8568 396 bp DNA linear VRL 15-APR-2005  
DEFINITION Human immunodeficiency virus type 2 pol gene, isolate b1053.  
ACCESSION AJ008568  
VERSION AJ008568.1 GI:3286892  
KEYWORDS DNA polymerase; pol gene.  
SOURCE Human immunodeficiency virus 2 (HIV-2)  
ORGANISM Human immunodeficiency virus 2

Viruses; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.

1 Xiang,Z., Ariyoshi,K., Wilkins,A., Dias,F., Whittle,H. and

Breuer,J.  
HIV type 2 pathogenicity is not related to subtype in rural Guinea

Bisseau  
AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)

Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F.,

Whittle,H. and Breuer,J.  
Correlation of HIV-2 genotype with progression to AIDS in vivo

Unpublished  
3 (bases 1 to 396)

Xiang,Z.  
Direct Submission

Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,  
Department of Virology, Queen Mary Westfield, Turner Street,

London, E1 2AD, UK

Db 194 CAAGTAAATGCAGAAATAGG 213

## RESULT 12

## LOCUS

## HIVAJ8582

## DEFINITION

## HIVAJ8582

## ACCESSION

## AJ008582

## VERSION

## AJ008582.1

## KEYWORDS

## DNA polymerase; pol gene.

## SOURCE

## Human immunodeficiency virus 2 (HIV-2)

## ORGANISM

## Human immunodeficiency virus 2

Viruses; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.

## REFERENCE

## 1

## Xiang,Z., Ariyoshi,K., Wilkins,A., Dias,F., Whittle,H. and

Breuer,J.  
HIV type 2 pathogenicity is not related to subtype in rural GuineaBisseau  
AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)

## JOURNAL

## PUBMED

## 9100992

## REFERENCE

## 2

## Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F.,

Whittle,H. and Breuer,J.  
Correlation of HIV-2 genotype with progression to AIDS in vivoUnpublished  
3 (bases 1 to 396)Xiang,Z.  
Direct SubmissionSubmitted (21-MAY-1998) Xiang Z., London Hospital Medical College,  
Department of Virology, Queen Mary Westfield, Turner Street,

## London, E1 2AD, UK

## FEATURES

## Location/Qualifiers

## 1..396

## /organism="Human immunodeficiency virus 2"

## /mol\_type="genomic DNA"

## /strain="A"

## /isolate="b1093"

## /db\_xref="taxon:11709"

## &lt;1..&gt;396

## /gene="pol"

## &lt;1..&gt;396

## /gene="pol"

## /codon\_start=2

## /product="DNA polymerase"

## /protein\_id="CAA08046.1"

## /db\_xref="GI:3286921"

## /db\_xref="GOA:Q9YQK6"

## /db\_xref="InterPro:IPR001584"

## /db\_xref="UniProt/TREMBL:Q9YQK6"

/translation="DHLVSGIRQVLFLEKIEPAQBEHEKYHNNKELSHKFGFLPTLV  
ARQVNTCAQCKQKGAHQVNAEIGTWQMDCTHLEKGIIVAVHVASGFEAEVVP  
QESGRQTALFLKLASRWPTIHLTDNGA"

## ORIGIN

## Query Match

## Best Local Similarity

## Matches

## 19; Conservative

## 0; Mismatches

## 1; Indels

## 0; Gaps

## 0;

## Qy 1 CAAGTAAATGCAGAAACAGG 20

## Db 194 CAAGTAAATGCAGAAATAGG 213

## RESULT 13

## LOCUS

## HIVAJ8589

## DEFINITION

## HIVAJ8589

## ACCESSION

## AJ008589

## VERSION

## AJ008589.1

## KEYWORDS

## DNA polymerase; pol gene.

## SOURCE

## Human immunodeficiency virus 2 (HIV-2)



source  
1. .396  
/organism="Human immunodeficiency virus 2"  
/mol\_type="genomic DNA"  
/strain="A"  
/isolates="b1244"  
/db\_xref="taxon:11709"  
<1..>396  
/genes="pol"  
<1..>396  
/gene="pol"  
/codon\_start=2  
/product="DNA polymerase"  
/protein\_id="CAA08106.1"  
/db\_xref="GI:3287041"  
/db\_xref="GOA:O9YOH6"  
/db\_xref="InterPro:IPR001584"  
/db\_xref="UniProt/TREMBL:Q9YOH6"  
/translation="DHLVSGQIRQVLFLEKIEPAQOEHEKYHSNVKELSHKFGLPKLV  
ARQIVNTCAQCOQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFIEAEVIP  
QESGRQTALFLKLASRWPIITHLHTDNGA"

## ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 396;  
Best Local Similarity 95.0%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20  
|||||  
Db 194 CAAGTAAATGCAGAAATAGG 213

RESULT 16  
HIVAJ8646  
LOCUS  
DEFINITION  
HIV type 2 pathogenicity is not related to subtype in rural Guinea  
ACCESSION  
AJ008646  
VERSION  
AJ008646.1 GI:3287048  
KEYWORDS  
DNA polymerase; pol gene.  
SOURCE  
Human immunodeficiency virus 2 (HIV-2)  
ORGANISM  
Human immunodeficiency virus 2  
VIRUSES; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.  
REFERENCE  
1 Xiang, Z., Ariyoshi, K., Wilkins, A., Dias, F., Whittle, H. and  
Breuer, J.  
AUTHORS  
HIV type 2 pathogenicity is not related to subtype in rural Guinea  
Bissau  
JOURNAL  
AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)  
PUBMED  
9100992  
REFERENCE  
2  
Grassly, N., Xiang, Z., Ariyoshi, K., Aaby, P., Jensen, H., Dias, F.,  
Whittle, H. and Breuer, J.  
AUTHORS  
Correlation of HIV-2 genotype with progression to AID in vivo  
Unpublished  
JOURNAL  
Xiang, Z.  
REFERENCE  
3 (bases 1 to 396)  
Direct Submission  
TITLE  
Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,  
Department of Virology, Queen Mary Westfield, Turner Street,  
London, E1 2AD, UK

FEATURES  
source  
1. .396  
/organism="Human immunodeficiency virus 2"  
/mol\_type="genomic DNA"  
/strain="A"  
/isolates="b1253"  
/db\_xref="taxon:11709"  
<1..>396  
/gene="pol"  
<1..>396  
/gene="pol"  
/codon\_start=2

gene  
CDS

/product="DNA polymerase"  
/protein\_id="CAA08110.1"  
/db\_xref="GI:3287049"  
/db\_xref="GOA:O77C38"  
/db\_xref="InterPro:IPR001584"  
/db\_xref="InterPro:IPR003308"  
/db\_xref="UniProt/TREMBL:Q77C38"  
/translation="DHLVSGQIRQVLFLEKIEPAQOEHEKYHSNVRELSHKFGLPKLV  
ARQIVNTCAQCOQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFIEAEVIP  
QESGRQTALFLKLASRWPIITHLHTDNGA"

## ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 396;  
Best Local Similarity 95.0%; Pred. No. 4.9e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20  
|||||  
Db 194 CAAGTAAATGCAGAAATAGG 213

## RESULT 17

HIVAJ8656  
LOCUS  
DEFINITION  
HIV type 2 pathogenicity is not related to subtype in rural Guinea  
ACCESSION  
AJ008656  
VERSION  
AJ008656.1 GI:3287068  
KEYWORDS  
DNA polymerase; pol gene.  
SOURCE  
Human immunodeficiency virus 2 (HIV-2)  
ORGANISM  
Human immunodeficiency virus 2  
VIRUSES; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.  
REFERENCE  
1 Xiang, Z., Ariyoshi, K., Wilkins, A., Dias, F., Whittle, H. and  
Breuer, J.  
AUTHORS  
HIV type 2 pathogenicity is not related to subtype in rural Guinea  
Bissau  
JOURNAL  
AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)  
PUBMED  
9100992  
REFERENCE  
2  
Grassly, N., Xiang, Z., Ariyoshi, K., Aaby, P., Jensen, H., Dias, F.,  
Whittle, H. and Breuer, J.  
AUTHORS  
Correlation of HIV-2 genotype with progression to AID in vivo  
Unpublished  
JOURNAL  
Xiang, Z.  
REFERENCE  
3 (bases 1 to 396)  
Direct Submission  
TITLE  
Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College,  
Department of Virology, Queen Mary Westfield, Turner Street,  
London, E1 2AD, UK

## FEATURES

source  
1. .396  
/organism="Human immunodeficiency virus 2"  
/mol\_type="genomic DNA"  
/strain="A"  
/isolates="b1285"  
/db\_xref="taxon:11709"  
<1..>396  
/genes="pol"  
<1..>396  
/genes="pol"  
/codon\_start=2  
/product="DNA polymerase"  
/protein\_id="CAA08120.1"  
/db\_xref="GI:3287069"  
/db\_xref="GOA:O9YJ02"  
/db\_xref="InterPro:IPR001584"  
/db\_xref="InterPro:IPR003308"  
/db\_xref="UniProt/TREMBL:Q9YJ02"  
/translation="DHLVSGQIRQVLFLEKIEPAQOEHEKYHSNVRELSHKFGLPKLV  
ARQIVNTCAQCOQGEAIGHQVNAEIGTWQMDCTHLEGGKIIIVAVHVASGFIEAEVIP  
QESGRQTALFLKLASRWPIITHLHTDNGA"

gene  
CDS

## ORIGIN

```

Query Match      92.0%; Score 18.4; DB 13; Length 396;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCGAAGAACAGG 20
    |||||
Db 194 CAAGTAAATGCGAAGAACAGG 213

RESULT 18
LOCUS HIVAJ8543 398 bp DNA linear VRL 15-APR-2005
DEFINITION Human immunodeficiency virus type 2 pol gene, isolate b1010.
ACCESSION AJ008543
VERSION AJ008543.1 GI:3286842
KEYWORDS DNA polymerase; pol gene.
SOURCE Human immunodeficiency virus 2 (HIV-2)
ORGANISM Human immunodeficiency virus 2
REFERENCE 1
AUTHORS Xiang,Z., Ariyoshi,K., Wilkins,A., Dias,F., Whittle,H. and Breuer,J.
TITLE HIV type 2 pathogenicity is not related to subtype in rural Guinea Bissau
JOURNAL AIDS Res. Hum. Retroviruses 13 (6), 501-505 (1997)
PUBMED 9100992
REFERENCE 2
AUTHORS Grassly,N., Xiang,Z., Ariyoshi,K., Aaby,P., Jensen,H., Dias,F., Whittle,H. and Breuer,J.
TITLE Correlation of HIV-2 genotype with progression to AID in vivo
JOURNAL Unpublished
AUTHORS Xiang,Z.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1998) Xiang Z., London Hospital Medical College, Department of Virology, Queen Mary Westfield, Turner Street, London, E1 2AD, UK
FEATURES
source
1..398
/organism="Human immunodeficiency virus 2"
/mol_type="genomic DNA"
/strain="A"
/isolate="b1010"
/db_xref="taxon:11709"
<1..>398
/gene="pol"
<1..>398
/gene="pol"
/codon_start=1
/product="DNA polymerase"
/protein_id="CAA08007.1"
/db_xref="GI:3286843"
/db_xref="GOA:O91095"
/db_xref="InterPro:IPR001584"
/db_xref="UniProt:IPR003308"
/db_xref="UniProt/TREMBL:O91095"
/translatation="RPFSGVRASDKYCSWKESPLREBEKHYSHNVKLSHKFGLPKL
VARQIVNTCAQCQCKGEATHGOVNAEIGTWQMDCTHLEKGIIVAVHVASGFIEAIV
PQESGRQTALFLKLASRPITHLTDNGA"

ORIGIN
Query Match      92.0%; Score 18.4; DB 13; Length 398;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCGAAGAACAGG 20
    |||||
Db 196 CAAGTAAATGCGAAGAACAGG 215

RESULT 19

```

```

AF395561
LOCUS AF395561 445 bp DNA linear VRL 05-SEP-2002
DEFINITION HIV-2 isolate HIV-2-BBI09 from Cote d'Ivoire pol polyprotein (pol)
gene, partial cds.
ACCESSION AF395561
VERSION AF395561.1 GI:18643070
KEYWORDS
SOURCE Human immunodeficiency virus 2 (HIV-2)
ORGANISM Human immunodeficiency virus 2
REFERENCE 1
AUTHORS Masclotra,S., Yang,C., Pieniazek,D., Thomas,C., Owen,S.M., McClure,H.M. and Lal,R.B.
TITLE Detection of Simian Immunodeficiency Virus in Diverse Species and of Human Immunodeficiency Virus Type 2 by Using Consensus Primers within the pol Region
JOURNAL J. Clin. Microbiol. 40 (9), 3167-3171 (2002)
PUBMED 12202548
REFERENCE 2
AUTHORS Masclotra,S., Yang,C., Pieniazek,D., Thomas,C., Owen,S.M., McClure,H.M. and Lal,R.B.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2001) HIV Immunology and Diagnostic Branch, Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA
FEATURES
source
1..445
/organism="Human immunodeficiency virus 2"
/mol_type="genomic DNA"
/isolate="HIV-2-BBI09"
/db_xref="taxon:11709"
/country="Cote d'Ivoire"
/note="isolated from PBMCs"
<1..>445
/gene="pol"
<1..>445
/gene="pol"
/codon_start=1
/product="pol polyprotein"
/protein_id="AAL76221.1"
/db_xref="GI:18643071"
/translations="CQKGEAIGHGVNAEIGVQMDCTHLEKGIIVAVHVASGFIEA
EVIFQESGRQTALFLKLASRPITHLTDNGSNFTSDEVKMAVWAMGIEQSGVPYN
PQSQGVVEMNHHLKQISRTREQANTVETIIVLMAVHCHMNFRRRE"

ORIGIN
Query Match      92.0%; Score 18.4; DB 13; Length 445;
Best Local Similarity 95.0%; Pred. No. 4.9e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCGAAGAACAGG 20
    |||||
Db 31 CAAGTAAATGCGAAGAACAGG 50

RESULT 20
HIVFAP
LOCUS HIVFAP 708 bp ss-RNA linear VRL 08-DEC-1994
DEFINITION Human immunodeficiency virus type 2 (clone 1) pol gene, partial cds.
ACCESSION L33089
VERSION L33089.1 GI:533792
KEYWORDS pol gene.
SOURCE Human immunodeficiency virus 2 (HIV-2)
ORGANISM Human immunodeficiency virus 2
REFERENCE 1
AUTHORS Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae; Lentivirus; Primate lentivirus group.
1 (bases 1 to 708)
Gao,F., Yue,L., Robertson,D.L., Hill,S.C., Hui,H., Biggar,R.J., Neequaye,A.E., Whelan,T.M., Ho,D.D., Shaw,G.M., Sharp,P.M. and Hahn,B.H.

```



**TITLE** Genetic diversity of human immunodeficiency virus type 2: evidence for distinct sequence subtypes with differences in virus biology

**JOURNAL** J. Virol. 68 (11), 7433-7447 (1994)

**PUBMED** 7933127

**COMMENT** On Sep 9, 1994 this sequence version replaced gi:532151. Original source text: Human immunodeficiency virus type 2 (individual isolate FA) RNA.

**FEATURES**

source

1..708  
/organism="Human immunodeficiency virus 2"  
/mol\_type="genomic RNA"  
/isolate="FA"  
/db\_xref="taxon:11709"  
/clone="1"

1..708  
/gene="pol"  
<1..>708  
/gene="pol"  
/codon\_start=2  
/product="pol protein"  
/protein\_id="AA56939.1"  
/db\_xref="GI:533178"  
/translation="MGIVVQPTSENRIVNQIIEDMIKKEAIYVAVVPAHKGIGNQ EVDHLVSGIQRVFLIERIEPAQEEHEKHSNIKELTHKFGIPQLVARQIVNTCAQCQ QKEATHGQVNAEIGVQWMDCTHLEKGIIVAVHVASGFIEAKVIPQSERQTALFL KLASRPITLHTDNGSNFTSQEVKVAWVWIGIEQTFGVYPNPQSQGVVEAMNHLKN QISRIREQANTTETIVL"

**ORIGIN**

Query Match 92.0%; Score 18.4; DB 13; Length 708;  
Best Local Similarity 95.0%; Pred. No. 4.8e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
|||||  
Db 332 CAAGTAAATGCAGAAATAGG 351

**RESULT 21**

HIV2D194 9472 bp DNA linear VRL 28-MAR-2000

**LOCUS** Human immunodeficiency virus type 2 (HIV-2), complete proviral genome.

**DEFINITION** J04542

**ACCESSION** J04542.1 GI:325654

**VERSION** complete genome; core protein; env gene; gag gene; nef gene; pol gene; regulatory factor; rev gene; reverse transcriptase; tat gene; vif gene; vpr gene; vpx gene.

**KEYWORDS** Human immunodeficiency virus 2 (HIV-2)

**SOURCE** Human immunodeficiency virus 2 (HIV-2)

**ORGANISM** Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae; Lentivirus; Primate lentivirus group.

**REFERENCE** 1 (bases 1 to 9472)

**AUTHORS** Kuehnel,H., von Briesen,H., Dietrich,U., Adamski,M., Mix,D., Biesert,L., Kreutz,R., Immelmann,A., Henco,K., Meichaner,C., Andreesen,R., Gelderblom,H. and Ruebsamen-Waigmann,H. Unpublished

**JOURNAL** 2 (sites)

**REFERENCE** Kuehnel,H., von Briesen,H., Dietrich,U., Adamski,M., Mix,D., Biesert,L., Kreutz,R., Immelmann,A., Henco,K., Meichaner,C., Andreesen,R., Gelderblom,H. and Ruebsamen-Waigmann,H.

**AUTHORS** Molecular cloning of two west African human immunodeficiency virus type 2 isolates that replicate well in macrophages: a Gambian isolate, from a patient with neurologic acquired immunodeficiency syndrome, and a highly divergent Ghanian isolate

**JOURNAL** Proc. Natl. Acad. Sci. U.S.A. 86 (7), 2383-2387 (1989)

**PUBMED** 2467304

**COMMENT** The revision in reference [2] corrected a possible frameshift mutation in gag and a premature stop in env observed in the original copy of the sequence.  
[2] sites; for [1].  
The gag coding sequence appears to be distorted at the 5' end by a frameshift; the env cds has a premature stop codon. The 5' ends of

**FEATURES**

source

the pol and gag genes have not been determined.

Location/Qualifiers

1..9472  
/organism="Human immunodeficiency virus 2"  
/proviral  
/mol\_type="genomic DNA"  
/isolate="D194"  
/db\_xref="taxon:11709"  
1..299  
/note="5"  
/evidence=experimental  
1..173  
/standard\_name="R repeat"  
/evidence=experimental  
59..120  
/note="aggtagagcctgggtgtccctgtagactctcaccagtgcttgcccgga ctggcacagc in [2] revises ag in [1]"  
303..319  
/bound\_moiety="Lys-tRNA primer"  
/evidence=experimental  
547..2112  
/gene="gag"  
547..2112  
/gene="gag"  
/codon\_start=1  
/evidence=experimental  
/protein\_id="AAA76840.1"  
/db\_xref="GI:325657"  
/translation="MGARNSVLRGKKADELKVRLRPNRGKRYRLKLVVVAANELDRF GLAESLLSKEGCKILKVLPLVPTSGENKSLFNTVCVITCLHAEKVKDTTEAKK LAQRHLVABTGAETKMPNISRPTAPSGKGPNPQQAGGNIHVPLSPRTLNAAVKL VEEKFGAEVWFGFQALSEGCTPYDINQMLNCVGDHQAQWQIIRIINEEAAWDQAQ PTPGLPQLDRGSDIAGTTSTVDEQIOMYQPNPVPVGNIVRRWIGLQKCV RMYNTNLDVRQGPKEFSQSVDRFYKSLRAEQTDPAVKNMTTLLIQANPDCKL VLKGLMPTLEMLTACQGVGSPQKARAEALKALTAPAPIPAQAQORRAIRCM NCKEGHSAKQCRAPRRQSCWKQSGKSHIMANCPERQAGFLGPRGKQPRNPAQA PQGLIPTAPPIDPAVDLLEKYNQOQKQREQRPYKEVTEDLLHLEQETPRGATE DLLHLSLFGKQD"

939..940  
/gene="gag"  
/note="ag in [2] revises agg in [1]"  
<1716..>4937  
/gene="pol"  
<1716..>4937  
/gene="pol"  
/codon\_start=1  
/evidence=experimental  
/protein\_id="AAA76841.2"  
/db\_xref="GI:7335662"  
/translation="VLELWKGGLTGVSTPSTQKTGLLEVWQVTRHGKLPKGTGRPFR DGPTGKAAPQLPRGPSSSGADTNPNSSSGPGVEIYAAREKAERAEGETIQGGDG LTAPRAGRDPORGDRGLATPOPSLWKPVTAFIEDQPVVELLDTGADDSIVAGIEL GNYTPKIVGGIGGFINTKEYKNVEIKLVKRVATIMGTDPITFGRNILATLWS LNLPAKLDPIKVTLPKGDGPRKQWPLTKEIKELKECEKMEREGOLEAPPTNP YNTPFAIKKDKKNRMLIDPRELNRVTDFTFEOIGPHPGAGLAKKXITVLVDGD AVPSIFLHEDFROYTATLPSVNNAPPEKRYVYKVPQWKGSPATFQVWROILEPF RKANPDVILIQYMDLILASDRTLGDKHVLQKELLNGLGFSTDEKFKQDPFQM MYELMPTKQKIQIOLPOKEIWTVDIQKLVGLNAAQIYPGIKTKHLCKLIRGM TLTVEQVTELAELAELENKIILSQEGSYQEESELEATVIKSQDNQWAKIHOQE RVLKVGKAKIKNTHNGVRLLAQVQKIGKALVETVETGDCNRSKQEGKAGYVTDG YQVTVWPFWDVFSTPTLVLTLNVLGDPDIPGTFTFYDGSCNRKQEGKAGYVTDG RDVRVLEOTSNQAELEAFAMALADSPKPNIVDSQYVMGIVAGOTSENRINVQ IEDMTKKBAVYVAVWPAHKGIGNOBVDHLVSGQIRQVLFLEKIEPAQEBHEKYSN IKELTHKFGIPQLVARQIVNTCAQCQKGEAIGHQVNAEIGVQWMDCTHLEKGIIVA HVASGFIEAIVIPQSGVROTALFLKLASRWITLHTDNGSNFTSQEVKVAWVWIG IEQSGFVYPNPQSQGVVEAMNHLKNQISRIREQANTTETIVLMAVCMFKRRGGIG DMTFABRLINMITTEQIEQFLQRKNSNFKKFYVYREGDQLWKGPGELLWKGQGA VI VKYGADIKVVPKRKAKIIRDYGGRQELDSSSHLEGAREDEGEVA"

4867..5514  
/gene="vif"  
4867..5514  
/gene="vif"  
/codon\_start=1

```
/evidence=experimental
/db_xref="GI:325659"
/translation="MEEGKNWLVPTWVRPGRMERWHSLVKHLKYRTKDLBEVRVYP
HKGAWMTCSVIFPLEGESHLEIQAYWNLTPKGMWSSHSVRLTWYTKFPWTDVTP
DCADSLIHSTFAGEVRRIRGEKLLSCCNPQAHKAQVPSLOYLALVVOQNG
RPQRGAARKQWRDRHWRGLRVARQDYSRLKQGGSESPAPRAHFPQVAKVLGILA"
gene
5342. .5680
/genes="vpx"
CDS
5342. .5680
/genes="vpx"
/codon_start=1
/evidence=experimental
/protein_id="AA76843.1"
/db_xref="GI:325660"
/translation="WADPRVRVPPGNSGEETIGEAFWLDRTIEALNREAVNHLPREL
IFQVQRQSWAYWHDEQGMSTYTRYLCIMQKAVYIHFKKGCTCLGRGHGPGWRRPG
PPPPPPGLV"
gene
5680. .5997
/genes="vpr"
CDS
5680. .5997
/genes="vpr"
/codon_start=1
/evidence=experimental
/protein_id="AAA76844.1"
/db_xref="GI:325661"
/translation="WTEAPTEPPDPGTPRRELGSTWVETLKEIKERALKHFPDCLL
IALGNYIYNRHGDTLEGARELIRVLQALFVHIRAGCDRSRKGQTRRAPCPAAPTFR
GMH"
gene
5843. .8371
/genes="tat"
CDS
Join(5843. .6138,8275. .8371)
/genes="tat"
/codon_start=1
/evidence=experimental
/protein_id="AAA76845.1"
/db_xref="GI:325655"
/translation="WETPLKEPESLESYNEPSSCTSRDVTVAQERAKQGBELLAQLH
RPLEATSCYCKQSCYHCQJCLFKKGLGIWYARQGRRTTRPKTKTHPPPSAKSIS
TRTGSQTPKQKQTPETTVWSACGLH"
gene
6069. .8516
/genes="rev1"
CDS
Join(6069. .6138,8275. .8516)
/genes="rev1"
/codon_start=1
/evidence=experimental
/protein_id="AAA76846.1"
/db_xref="GI:325656"
/translation="WRDRADERGLOEKLRLIRLLHQINPYPHGQGTASORNNRRRRR
RQWRLVALATKLHTIPDPPTSDPLDRAIQQLGLTIQELPDPPTDLPESNSQGLAE
T"
gene
6145. .8700
/genes="env"
CDS
6145. .8700
/genes="env"
/codon_start=1
/evidence=experimental
/protein_id="AAA76847.1"
/db_xref="GI:325662"
/translation="MEPGRNQLLVAILLTSACLIYCKQVTVFYGIPAWRNASIPLFC
ATKNRDWTGTOCLDNDYQBITNLNVEAPDAWNTVEQALIEDVWRLFETSIKPCV
KLTPUCVANNCNITSGTATPSPFNITIDENSTCIGNNCTGLGKEEVEECFNNGT
LEQDKRKNDAWYRDVCDKNTGTCTYMRHCNTSVIKESCDSKHWDAMKFRYCAP
PGFALLRNDNTYSGFEPKCSKVAASCTRMETQSTWFGFNTRAENRTIYWHGK
DNRTIISLKNYLNTHCKRPGNKTVPTILMSGRFRFSRPVYNKKPGQAWCFQGNW
TEAMREVKTAKHPRYGCTNDTKGIFKPGIGSDPEVTYMMTCRGEFLYCNMTWF
LNWENKNTQTHGNAPCHIRQIINTWVKVGNVYLPREBELTCNSTVTSIIANIDS
DGNQNTIFSABEADYRLGDKLIEVTPIPPAPTREKYSAPVRNKRGVFVLG
LGLFAGSAGMGSALTSAQSTKTLGAGIVOOQQLLDVVRKQOEMRLTVWGTKQLQ
ARVTAIKDQKQAGLNSWGCARFQVCHTTPWVNDSLTPWNWNTWQEWKRVHYLE
ANISOLEAQIQQSKNMYLOKLSWDVFGNFDLTSWKYIQYGVTVVIGLIRI
ATYIVQLLSRLKGRVPFSSPGVYLQIHIHTRGQPANBEEDAGDSDGFLGWPW
PLNYIQFLHLLTRLLTGLYNSCRGLLSKNSPTRLLISQSLTAIRDWLRLKAYLYQY
```

```
variation
6940. .6943
/genes="rev1"
/notes="gaaa in [2] revises ga in [1]"
6971. .6973
/genes="rev1"
/notes="agg in [2] revises ag in [1]"
6999. .7001
/genes="rev1"
/notes="caa in [2] revises ca in [1]"
7682. .7685
/genes="rev1"
/notes="cgac in [2] revises cagc in [1]"
7707. .7709
/genes="rev1"
/notes="cgc in [2] revises cnc in [1]"
8534. .9307
/genes="nef"
8534. .9307
/genes="nef"
/codon_start=1
Query Match 92.0%; Score 18.4; DB 13; Length 9472;
Best Local Similarity 95.0%; Pred. NO. 4.3e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAAGTAATGCAGAAACAGG 20
|||||
Db 4212 CAAGTAATGCAGAAATAGG 4231
|||||

RESULT 22
HIVD194
LOCUS HIVD194 9472 bp DNA linear VRL 18-APR-2005
DEFINITION Human immunodeficiency virus type 2 strain D194 proviral genome.
ACCESSION X52223
VERSION X52223.1 GI:60155
KEYWORDS env gene; envelope glycoprotein; gag gene; integrase; long terminal
repeat; nef gene; pol gene; polymerase; reverse transcriptase;
RNase H; tat gene; vif gene; vif gene; vpx gene.
SOURCE Human immunodeficiency virus 2
ORGANISM Human immunodeficiency virus 2
Viruses; Retro-transcribing viruses; Retroviridae;
Orthoretrovirinae; Lentivirus; Primate lentivirus group.
REFERENCE
1
AUTHORS Kuehn,H., Kreutz,R. and Rubsamen-Waigmann,H.
TITLE Nucleotide sequence of HIV-2D194, an isolate from a Gambian case of
'neuro-AIDS', which showed excellent growth in macrophages
JOURNAL Nucleic Acids Res. 18 (20), 6142 (1990)
PUBMED 2235509
REFERENCE
2 (bases 1 to 9472)
AUTHORS Kuehn,H.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1990) Kuehn,H., Chemotherapeutisches
Forschungsinstitut, Georg-Speyer-Haus, Paul-Ehrlich-Strasse 42-44,
600 Frankfurt 70, F R G
COMMENT Virus DNA from a genomic library of infected cord blood
lymphocytes. HIV-2 strain D194 was isolated in 1987 from
peripheral blood of a Gambian male with neurological manifestations
of AIDS.
Ribosomal frame shift gives rise to pol gene product. See also
<J04542> for overlapping sequences.
FEATURES
source
1. .9472
/organism="Human immunodeficiency virus 2"
/mol_type="genomic DNA"
/strain="D194"
/db_xref="taxon:11709"
/cell_type="cord blood lymphocyte"
1. .173
/notes="R region"
174. .299
/notes="U5 region"
misc_feature
misc_feature
```



KEYWORDS	Human immunodeficiency virus 2 (HIV-2)	
SOURCE	Human immunodeficiency virus 2	
ORGANISM	Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae; Lentivirus; Primate lentivirus group.	
REFERENCE	1 (bases 1 to 9473)	
AUTHORS	Henco,K., von Briesen,H., Immelman,A., Kuehnel,H., Dietrich,U., Ruedsammen-Wagmann,H. and Adamski,M.	
TITLE	HIV-2 virus variants	
JOURNAL	Patent: EP 0347365-A 1 20-DEC-1989;	
DIAGN	Institut fuer molekularbiologische Diagnostik GmbH;	
CHEMOTHERAPEUTISCHES FORSCHUNGSINSTITUT GEORG-SPEYER-HAUS		
FEATURES	Location/Qualifiers	
source	1..9473	
ORIGIN	/organism="Human immunodeficiency virus 2"	
	/mol_type="unassigned DNA"	
	/db_xref="taxon:11709"	
Query Match	92.0%; Score 18.4; DB 6; Length 9473;	
Best Local Similarity	95.0%; Pred. No. 4.3e+02;	
Matches	19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Qy	1 CAAGTAAATGCAGAACAGG 20	gene
Db	4213 CAAGTAAATGCAGAAATAGG 4232	CDS
RESULT 24		
HIV2U38293		
LOCUS	HIV2U38293 10312 bp DNA linear VRL 20-FEB-1997	
DEFINITION	Human immunodeficiency virus type 2, complete proviral genome.	
ACCESSION	U38293	
VERSION	U38293.1 GI:1845204	
KEYWORDS		
SOURCE	Human immunodeficiency virus 2 (HIV-2)	
ORGANISM	Viruses; Retro-transcribing viruses; Retroviridae; Orthoretrovirinae; Lentivirus; Primate lentivirus group.	
REFERENCE	1 (bases 1 to 10312)	
AUTHORS	Barnett,S.W.	
TITLE	Direct Submission	
JOURNAL	Submitted (11-OCT-1995) Barton Bryant, HIV Database, Los Alamos National Laboratory, Eniwetok St., Los Alamos, NM 87545, USA	
FEATURES	Location/Qualifiers	
source	1..10312	
ORIGIN	/organism="Human immunodeficiency virus 2"	
	/proviral	
	/mol_type="genomic DNA"	
	/db_xref="taxon:11709"	
	/note="similar to HIV2BEN, GenBank Accession Number M30502, and to HIV2D194, GenBank Accession Number J04542"	
gene	1103..2668	
CDS	1103..2668	
	/gene="gag"	
	/codon_start=1	
	/evidence=not_experimental	
	/product="gag polyprotein"	
	/protein_id="AAB47780.1"	
	/db_xref="GI:1845205"	
	/translation="MGARNSVLRGKKADELEKIRLIRPGKKYRLKHLIWAANEIDRFRLAENLSKEQCILVLDPLVPSGKNSLFTVCVWCLHAEKVKDTEAAKLIVORHLVAETGAEMPTNSRTPAPSGKGNFPVQAGNYPVHPLSPRTLNAMVKLVEEKFQVPGFOALSEGCTPYDINQMLNCVGGQAAQIIRIINEEAADAOHQVPIGPPLAGQLDRGSDIAGTSTVDQIOWMYQQQNPVPGNLYRBIQIGLOKCYRMYNTVNLVDVKQPKESFQSVDRFYSLRAEQADPAVKNWMTTLIVQANPCKLVLKGLGNPTLEEMLTACQGGCGKARLMAELKALTPPIPFPAQAQQRRTIRCWNCGRGSHAKQCRPRRCQGWCKGPHIMANCPDQAGFLMGSRGKPRFPVAQAPQGLIPTAPADPAADLLLEKYLQOQRKQREPRPYKEVTEDLLHLEQETPREATEDLLHLNLSLFGKQD"	
gene	2863..5493	
	/gene="pol"	
	/evidence=not_experimental	
	/product="pol polyprotein"	
	/protein_id="AAB47781.1"	
	/db_xref="GI:1845206"	
	/translation="MTGDTPINIFGRNLTALGMSLNPVARIPIKITLKPGDKGPRLKQPLTKREKVEALKEICERMEKEGQLEAPPTNPYPTPTFAIKKKQKRWMLIDFR ELNRVTQFTFETIQIGIPHPAGLAKKRITVLVDGDAVFSIPLYEDFRYATPLPSVN NVRGKRYIYKVLPGQWKGSPAIQYTMROQLLEPERKANPDVILIYQMDLIIASDRT GLBHDKVVLQKELLNGLGFSTPEEKFOKDPFQWGYELMTKWKLOKIQLOPKERTW TVNDIQKLVGLNAAQIYPGIKTKHLCLRLRGKMTLTEEVQWTELAEAEENRIIL DQEGHYHQEKELEATI QKSQDNQWYKIQHEKILKVGKYAKIKNTHNTNGVRLLA QVQKIGKEALVIWGRIPKPHLPVERETQWMDNYSQVWIPWDFVSTPLVRLTF NLVGDPIPTGTFYTDGSCNRQSKGAGYVTDGRDKRVLEQTNQQAELAPAMA LADSPKVNIIIVDSQYVYGIVASQPTSESRIVNQIIEDMIKKEAVYVAMTPAHKIG GNOEVDHLVSGQIRQVLFLEKI EIPAQBEHEKYSNIKELIHKFGI POLVARQIVNTCA CQOQGEAIIHQVNAEIGVMQMDCTHLEGKIIIVAVHVASGFI EAEVTPQSGRQAL FLKLASRPITHLHTDNGPNTSQEVKMAWVWYIGSQSGVFPYNPQSGQGVAMNHH LKNQISRIREQANTVETIVLMAVHCMPKRRGGIDMTPAERLINMITTEQEIQLQR KNSNFKFQVYTYREGRDQLWKPGGELLWKGEAVIVKVGADIKVVPKRKAKIIRDYGG ROELDSSPHLEGAREDEGEVA"	
gene	5423..6070	
	/gene="vif"	
	5423..6070	
	/gene="vif"	
	/codon_start=1	
	/evidence=not_experimental	
	/product="vif protein"	
	/protein_id="AAB47782.1"	
	/db_xref="GI:1845207"	
	/translation="MEEGKSWIVVPTWRVPGMRWHSLSVKYIKRTKDLGVRVYPH HKVGMWMTCSRVPFPLKGESHLQIAYMNLTPKGLSSHSVLTWTYKEFTDVTIP DCADILIHSTYFCCFAAGEVRAIRGKLLSCCNYPQAHKAQVPSLOYLALVYVQND RPORKGPARKOWRRDRHWRGLRVARQDYRSLEQRGKPPAPRAHPFGVAKVLEILA"	
gene	5898..6236	
	/gene="vpx"	
	5898..6236	
	/gene="vpx"	
	/codon_start=1	
	/evidence=not_experimental	
	/product="vpx protein"	
	/protein_id="AAB47783.1"	
	/db_xref="GI:1845208"	
	/translation="MTDPFRVRVPPGNSGEETIGFAEWLDRITI EALNREAVNHLPREL IFQVWQSRWRYWHDEQGLSSSYTKRYVLCMQKAMFIHLKKGCTCLGGGHPGGWRPG PPPPPGLV"	
gene	6341..8911	
	/gene="tat"	
	Join(6341..6636,8815..8911)	
	/gene="tat"	
	/codon_start=1	
	/product="tat protein"	
	/protein_id="AAB47786.1"	
	/db_xref="GI:1845211"	
	/translation="METPLKAPSSLSKNFSPCTSEWVAQAQLAKQGBELLAQLHL RPLACTNPYCKKCSFCLCKGLGIWYVRKGRRTPKTKTTPPTPPDKSIS IQTQDSRPTKKQKETPTPTVATTSLGR"	
gene	6567..9056	
	/gene="rev"	
	Join(6567..6636,8815..9056)	
	/gene="rev"	
	/codon_start=1	
	/product="rev protein"	
	/protein_id="AAB47787.1"	
	/db_xref="GI:1845212"	
	/translation="MCKADEEELRELRLRLRLHQTNPYPYRPGTAGQRNRRRRR RQWLRLALADKLHTADPPADSLDWAHQHQLTIRELPDPTDLDPESDSNQGIAE T"	
gene	6643..9240	
	/gene="env"	
	6643..9240	

FEATURES	source	[2]	9132 to 10359 were from MK6.
Location/Qualifiers	1..10359		
organism="Human immunodeficiency virus 2"			
proviral			
mol_type="genomic DNA"			
isolate="BEN"			
db_xref="taxon:11709"			
clone="MK[2,6]"			
LTR	1..855		
primer_bind	859..875		
CDS	1103..2668		
	/note="gag polyprotein"		
	/codon_start=1		
	/protein_id="AAB00736.1"		
	/db_xref="GI:325648"		
	/translat="MGARNSVLRGKKADELEKVRLRPGGKKYRUKLHIWAAANELDKF		
	GLAESLLESKEGCKILRVLDPLVPTSGENLSKSFNTVCVWCLHAEKVKDTEAAK		
	LAORHLVAETGTAEKMPNTRTPAPPSGKRGYVQQAQGNVHVPLSPRTLNAYVKL		
	VEKKFGAEVVPQALSEGCTPYDINOLACVGDHQAQMOIIRIINEEADMDSOH		
	PIPGFLPAQLRDPRGSDIAGTSTVDSQIQMWRPQNPVPVGNVTVYRWIOLGLOKV		
	RKNPTNILDIKQGPFPFSYVDVRFYKSLRAEQTDPAVNMGTITLLIQNANPCKL		
	VLKGLNMPTLEMLTACQGVGPGQKARLMAEALKEAMGSPSPFAAAQKRAIRYWA		
	NCGEGHSAROCRAPRQGCWKCGKPGHIMANCPERQAGFLGPRGKKRPNFPVTOA		
	PQGLIPTAPPADPAALLERYMQQGRKQREQRPYKEVTEDLLHLEQRETPHREBTE		
	DLLHLNLSLFGKQ"		
gene	2863..5754		
	/gene="pol"		
CDS	2863..5754		
	/note="pol"		
	/note="First start codon located at 2863.; ORF bordered by		
	stop codons at 5' (bp 2323) and 3' (bp 5754) ends."		
	/codon_start=1		
	/product="pol polyprotein"		
	/protein_id="AAB00737.1"		
	/db_xref="GI:1332356"		
	/translat="MTGDTPINIFGRNLTALGMSLNPVAKIEPIKVTLPKPGKQGR		
	LKQWPLTEKIEALKEICEKMEKEGLEEAPNTPVNTPTFAIKKKDKNWRMLIDPR		
	ELNKVDTFTEIQIGIPHPAGLAKKRISILDVGDAYSIPLEDPROVYATATLPAVN		
	NMEPKRYIYKVLPGQWKGSPAIFYTWRQVLEPPFRKANPDVILIQYMODILLASRT		
	GLEHDKVVLQKLINGLGFSTDEKQFPFQWMCGLMPTKWLKQKLQPLQKDIW		
	TWINDIKLVNAAQIYSGIKTKHLIRIGKMTLVEOVKWLAEALBENKIL		
	SOEORGYYQREKELEATIQKSGHQWYKIHOBEEKLVKGVAKIKNTHTNGVRLLA		
	QVQKIGKALVINGRIKPFHLPVERETWEQWMDNYQVWTIPENDFVSTPLVRLTF		
	NLVGDPIGAETFFYTDGSCNROSKEGKAGVTDGRKDKVKVLEQTNQOAELEVFMA		
	LADSGPKVNIIVDSQYVNGIVAGQTESNRINQIIEBIMIKKEAVYVAVWPAHKGIG		
	GNQEVHLVDSQIRQVLFLEKIEPAQEHEKHSIKELTHKPGIPLLVARQIVNSCA		
	OCQKGEAIIHGQVNAEIGVQMDYTHLEKIIIVAVHVASGPIEAEVYIQESGRQAL		
	FLKLASRPITLHTDNGPNTSOEVKVAWVGIEQSGFVYPNPSQGVVVEAMNHH		
	LKNQISIREQANTTETIVLMAVHCNFKRGGIGDMPAERLINMTTTEQIQIQLQR		
	KNSNPNFOVYVYREGRDQLWKGFLMHGEGAVIKVGTIDIKVPRRAKAIIRDYGG		
	ROELDSPHLEGAREDMACPCQVPTIQNKPRGALCSPPQGGMGMDVLDQGN1PT		
	TRKGSRTTGILEPNTRKRMALLSCSKINLVYRKVLDRCYPRLCRHPNT"		
CDS	5423..6070		
	/note="vif protein"		
	/codon_start=1		
	/product="vif protein"		
	/protein_id="AAB00738.1"		
	/db_xref="GI:325649"		
	/translat="MEEDRNWIVWPTVRPMKEMHALVKVLYKTKYTKDLEEVRYVPH		
	HKVGAWMTCSRVIPLOGKSHLEIOAYWNLTPEKWLSSHAVRLTWYTEKFWDTVP		
	DCADILLHSYTFSCETAGEVRAIRGEKLLSCCNYPQAKQVPSLYLALVWVQOND		
	RPOKGTARKQWRDRHRLVAREDRHRSKOGGSESPAPRAHPFGVAKVLEILA"		
	5898..6239		
	/note="vpx protein"		
	/codon_start=1		
	/product="vpx protein"		
	/protein_id="AAB00739.1"		
	/db_xref="GI:325650"		
	/translat="MTDPRVRVPPGNSGETTGAEFEWERTIEALNREAVNHLPREL		
	IFQVMORSRYWHDQSGMSASYTKRYLCLMOKAIFTHPRGCTCGEDMGREGLEDQ		
	GPPPPPPPGLV"		

Qy	1	CAAGTAAATGCAGAACAGG 20		92.0%; Score 18.4; DB 13; Length 10312;	Query Match
Db	4768	CAAGTAAATGCAGAAATAGG 4787		95.0%; Pred. No. 4.3e+02;	Best Local Similarity
				Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
ORIGIN					
<pre> /gene="env" /codon_start=1 /evidence=not_experimental /product="envelope polyprotein" /protein_id="AAB47784.1" /db_xref="GI:1845209" /translat="MEPGNQQLAVILLTSACLIYCKQYVTVFYGVVWRNASIPFLC ATKNRDTWGTIOCLPDNDYOEIPLNVEAFDAMNTVTEQAIEDVRLFTSIPKPCV KLTPLCVAMNCPVTGNNATAPKTAARPTNPSYLTII NESSTCVGADNCTGLGDE CHVCKENFTGLEQDKIKGYDTTWYSDDVDCSTNKTGNTTCYMRHCNTSVIKESCD KHYDSMFRICTPGYALLRCNDNYSGFAPNCPKVAASCTRMETQTSTWFGFNG TRAENRTIYHGRDNRTIISLNHYNLTMHCGRNKTIVVPTILMSGRHRSQAVIN KKPQAWKFWKNGWAMQEVKQTLAGHPRYKGTNDTSKINFKPGVSGSDPEVTYMT NCRGFFYCNWFLNMYVENRTSQORNYAPCHIRQIINTWHKVQYVYLPREGELT CNSVTYSIIANDTGNONTIFSAAVLYRLGLDYKLEITPIGFAPTSKERYSS APARNKRGVFLGLGLFATAGSANGAASLTLSAQSRLLAGIQQOQQLDVKRQO ELLRLTWGKTRQVTAIEKYLKQQLNSWGCTFRQVCHTVPVWVNDLSLTPRNN MTWQBEWQVLEANI SOSLEAQIQDEKNMYELOKLNMDVFNWFDLSWIKYIQ YGVYIVVGIIALRIAYVYQLSRFRKGRVPVFPSPGYLQIHIHTRDGPANETE GDAGASVGDVLPWPIYIQLLIHLTLTLLTGLYSICRDLLSANSPTRLLSQNLTAI RDWELKAAYLQYGCENIQEAFQARTARETLAGAWRLCKAVQIRIGLILAVPRII RQGAIIALL" 9074..9847 /gene="nef" 9074..9847 /gene="nef" /codon_start=1 /evidence=not_experimental /product="nef protein" /protein_id="AAB47785.1" /db_xref="GI:1845210" /translat="MGASGSKLSQSGRLRLLRARGGVYVROCASGSGSSQSQG GSGRQKSPSCQYRGDYNMTWRTPAMESOKESYKQNMDDVSDNDNQVGSVY TSRVPRLPNTYKLVDMSHPIKEKGLGEMFYSDRHRLLDLYLKEGSGVIDPQNYT HGPGRYPNCFGLMKLVNPSQAEEDETNCLMHPAQTSDYIDDIHGTLLVWRFSNM LAYEYKAFITYPEEFHGKSGLPEKWKAKLKARGIPYSE" </pre>					
gene					
CDS					

RESULT 25					
HIV2BEN					
LOCUS					
DEFINITION					
ACCESSION					
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
PUBLISHED					
COMMENT					

/gene="env"  
/codon\_start=1  
/evidence=not\_experimental  
/product="envelope polyprotein"  
/protein\_id="AAB47784.1"  
/db\_xref="GI:1845209"  
/translat="MEPGNQQLAVILLTSACLIYCKQYVTVFYGVVWRNASIPFLC  
ATKNRDTWGTIOCLPDNDYOEIPLNVEAFDAMNTVTEQAIEDVRLFTSIPKPCV  
KLTPLCVAMNCPVTGNNATAPKTAARPTNPSYLTII NESSTCVGADNCTGLGDE  
CHVCKENFTGLEQDKIKGYDTTWYSDDVDCSTNKTGNTTCYMRHCNTSVIKESCD  
KHYDSMFRICTPGYALLRCNDNYSGFAPNCPKVAASCTRMETQTSTWFGFNG  
TRAENRTIYHGRDNRTIISLNHYNLTMHCGRNKTIVVPTILMSGRHRSQAVIN  
KKPQAWKFWKNGWAMQEVKQTLAGHPRYKGTNDTSKINFKPGVSGSDPEVTYMT  
NCRGFFYCNWFLNMYVENRTSQORNYAPCHIRQIINTWHKVQYVYLPREGELT  
CNSVTYSIIANDTGNONTIFSAAVLYRLGLDYKLEITPIGFAPTSKERYSS  
APARNKRGVFLGLGLFATAGSANGAASLTLSAQSRLLAGIQQOQQLDVKRQO  
ELLRLTWGKTRQVTAIEKYLKQQLNSWGCTFRQVCHTVPVWVNDLSLTPRNN  
MTWQBEWQVLEANI SOSLEAQIQDEKNMYELOKLNMDVFNWFDLSWIKYIQ  
YGVYIVVGIIALRIAYVYQLSRFRKGRVPVFPSPGYLQIHIHTRDGPANETE  
GDAGASVGDVLPWPIYIQLLIHLTLTLLTGLYSICRDLLSANSPTRLLSQNLTAI  
RDWELKAAYLQYGCENIQEAFQARTARETLAGAWRLCKAVQIRIGLILAVPRII  
RQGAIIALL"  
9074..9847  
/gene="nef"  
9074..9847  
/gene="nef"  
/codon\_start=1  
/evidence=not\_experimental  
/product="nef protein"  
/protein\_id="AAB47785.1"  
/db\_xref="GI:1845210"  
/translat="MGASGSKLSQSGRLRLLRARGGVYVROCASGSGSSQSQG  
GSGRQKSPSCQYRGDYNMTWRTPAMESOKESYKQNMDDVSDNDNQVGSVY  
TSRVPRLPNTYKLVDMSHPIKEKGLGEMFYSDRHRLLDLYLKEGSGVIDPQNYT  
HGPGRYPNCFGLMKLVNPSQAEEDETNCLMHPAQTSDYIDDIHGTLLVWRFSNM  
LAYEYKAFITYPEEFHGKSGLPEKWKAKLKARGIPYSE"

92.0%; Score 18.4; DB 13; Length 10312;  
95.0%; Pred. No. 4.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAACAGG 20  
|||||  
Db 4768 CAAGTAAATGCAGAAATAGG 4787  
|||||

RESULT 25  
HIV2BEN 10359 bp DNA linear VRL 30-AUG-2002  
LOCUS Human immunodeficiency virus 2 isolate BEN, complete genome.  
DEFINITION M30502  
ACCESSION M30502.1 GI:1332355  
VERSION  
KEYWORDS  
SOURCE Human immunodeficiency virus 2 (HIV-2)  
ORGANISM Human immunodeficiency virus 2  
Viruses; Retro-transcribing viruses; Retroviridae;  
Orthoretrovirinae; Lentivirus; Primate lentivirus group.  
REFERENCE 2 (bases 1 to 10359)  
AUTHORS Kirschhoff,F., Jentsch,K., Bachmann,B., Stuke,A., Laloux,C.,  
Lueke,W., Stahl-Henning,C., Schneider,J., Nieselt,K., Eigen,M. and  
Hunsman,G.  
TITLE A novel proviral clone of HIV-2: biological and phylogenetic  
relationship to other primate immunodeficiency viruses  
JOURNAL Virology 177 (1), 305-311 (1990)  
PUBLISHED 2353457  
COMMENT On May 24, 1996 this sequence version replaced gi:325645.  
Kindly submitted prior to publication by Dr. F. Kirschhoff, DPZ,  
Gottengen, 0551/3851-0. HIV2BEN was isolated from a German AIDS  
patient (with predominantly neurological complications) who was  
probably infected in Mali (Klemm et al., J Neuro 235,304,1988). The  
sequence was determined from two clones: bases 1903 to 5057 and  
7293 to 9131 were from MK2 and bases 1 to 1902, 5058 to 7292 and

CDS 6239..6502  
/note="vpr protein"  
/codon\_start=1  
/protein\_id="AAB00740.1"  
/db\_xref="GI:325651"  
/translation="WTEAPTEPPEDGTPRDLGSDWVIETLREIKEBALRHFDPRL  
IALGVYINRHGDTLEGARELIKTLQALFVHFRAGCNRSIG"  
Join(6402..6697,8861..8957)  
/note="tat protein"  
/codon\_start=1  
/protein\_id="AAB00741.1"  
/db\_xref="GI:325646"  
/translation="METPLKAPESLKPYNPSSCTSERDVTQAQLAKOGBELLAQLH  
RPLECTNKCVCRCSCFQCFCFKGLGISFERGRRRPRKTKTSPSAPDKSIS  
TRTGDSPQTKSQKTSATVVTGCGQ"  
6402..6697  
/note="first expressed exon"  
/number=2  
Join(6628..6697,8861..9102)  
/note="rev protein"  
/codon\_start=1  
/protein\_id="AAB00742.1"  
/db\_xref="GI:325647"  
/translation="WSEADREGLOGKRLRLRLHQTNPYQPGCTASORNRERRR  
RQWLRLVALANKLCAVPDPDPLDRAIQHLQRLTIQELPDPPPTDLPESNQGLAE  
T"  
6628..6697  
/note="first expressed exon"  
/number=2  
6704..9286  
/note="env polyprotein"  
/codon\_start=1  
/protein\_id="AAB00743.1"  
/db\_xref="GI:325652"  
/translation="MEPCGNOLFVILLTSLACLVYCSQVTVFYGIPAWKNASIPLC  
ATKPRDWTGTOCLPDNDYOEIILNVTAEADANNVTTEQAVEDWHLFETSIKPCV  
KLTPUCVAMNCSRGVQNTTTPNRTSSITSRPPTSAASINETSNCIENTCAGLG  
BEMQCFNMGSDQKKRYKDTWYLEDVDCNTAGTCYNRHNTSIIKESCDHY  
WDAMFRYCAPGFCALLCNDNTNYSGFPCFKTKVVAASCTRMETQTSTWFGNGRA  
NRTYIYVHGRNRTIISLNKYNLTRCKRPGKNTLPITLMSGLVFHSOPINTRPR  
QAWCFGRGWRAMQEVKOTLVQHPRYKINDTKINFTKPGAGSDPEVAFWNTCRG  
RELYCNMTWELNWDKQTRNRYCHIKOILNTHKGVKNVLPREGELACESVTVS  
TIANDIDKNRTHNITSAEVAELRYRELGDYKLEIETPGAPTQDQRRVSSITVRN  
KRGVFLGFLATAGSAMGARSITLSAQSRLLIAGIVQQQQLLDVVKQQLMLRL  
TWGNTQARVTAIEKYLKHOALNSWGCAPRQVCHTTPVWVNDLSPDKWNTWQE  
WEKVRYLEANIQSLEAQIQOENMYELQKNSWDILGNWFDLTSWVKYIQYGVHI  
VVGIIALTAIVVOLLGRFRKGRPVFSSPPGYLQOIHKDRQOPANEGTEEDVG  
DSGYLWPKPIYVQFLIHLTRLLIGLYNICRDLKSNPTRLISQSLTAIRDMLR  
LKAALQVGCWEIQAFQAPARTTRETLAGAWGLWEAARRIGRILAVPRIRQAE  
LALL"  
8861..9102  
/number=3  
8861..8957  
/number=3  
9120..9893  
/note="nef protein"  
/codon\_start=1  
/protein\_id="AAB00744.1"  
/db\_xref="GI:325653"  
/translation="WGASGSKLSKHSRGLRERLARAGDYGKORDASGGEYSOFQE  
ESGRQNSPSCGQIQQGHYFNWPNRATERQDLYRQNMDDVDSDDDLIGVPV  
TPRVRRENTYKLAIDMSHFYKEKGLQGMFYRRRHRLDIYLEKBEGIIPDWQNT  
HGFVRYPMYFGLWKLVSLEAEDEANCLVHPAQTSRHHDEHGETTLVWQFDSM  
LAYNYKAFPLYPEEFHGKSGSLPEKEWKAKLKARGIPSE"  
9505..10359  
/note="3' LTR"  
10211..10216

polyA\_signal  
ORIGIN

Query Match 92.0%; Score 18.4; DB 13; Length 10359;  
Best Local Similarity 95.0%; Pred. No. 4.3e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20  
|||||  
Db 4768 CAACTAAATGCAGAAATAGG 4787  
|||||  
RESULT 26  
AY341245 37676 bp DNA linear PRI 22-JUL-2003  
LOCUS Homo sapiens chromosome condensation 1-like (CHC1L) gene, complete  
DEFINITION  
cde.  
ACCESSION AY341245  
VERSION AY341245.1 GI:32891796  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.  
REFERENCE  
AUTHORS 1 (bases 1 to 37676)  
Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W.,  
Miyamoto,K.E., Nguyen,C.P., Poel,C.L., Robertson,P.D.,  
Schackwitz,W.S., Sherwood,J.K., Witrak,D.A. and Nickerson,D.A.  
Direct Submission  
TITLE Submitted (11-JUL-2003) Genome Sciences, University of Washington,  
JOURNAL 1705 NE Pacific, Seattle, WA 98195, USA  
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome  
Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA  
(URL: http://egp.gs.washington.edu).  
FEATURES  
Location/Qualifiers  
source  
1..37676  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
misc\_feature  
1..346  
/note="Region not scanned for variation"  
repeat\_region  
32..51  
/rpt\_family="L1"  
/rpt\_type="dispersed"  
repeat\_region  
119..405  
/rpt\_family="L1"  
/rpt\_type="dispersed"  
variation  
557  
/frequency="0.01"  
/replace="c"  
725..36573  
/gene="CHC1L"  
join(725..833,3124..3219,3565..3629,9785..9940,  
10171..10321,12632..12798,13353..13511,13650..13757,  
14756..14898,22614..22804,23660..23786,25768..25907,  
29207..29337,35287..36573)  
/gene="CHC1L"  
/product="Chromosome condensation 1-like"  
variation  
852  
/gene="CHC1L"  
/frequency="0.01"  
/replace="a"  
983..1159  
/rpt\_family="MIR"  
/rpt\_type="dispersed"  
variation  
1085  
/gene="CHC1L"  
/frequency="0.01"  
/replace="g"  
1096  
/gene="CHC1L"  
/frequency="0.01"  
/replace="c"  
1327  
/gene="CHC1L"  
/frequency="0.04"  
/replace="c"  
1383  
/gene="CHC1L"  
variation





WPCOMMENT

Sequence split into 23 fragments LOCUS CR382125 Accession CR382125

Fragment Name	Begin	End
CR382125_00	1	110000
CR382125_01	100001	210000
CR382125_02	200001	310000
CR382125_03	300001	410000
CR382125_04	400001	510000
CR382125_05	500001	610000
CR382125_06	600001	710000
CR382125_07	700001	810000
CR382125_08	800001	910000
CR382125_09	900001	1010000
CR382125_10	1000001	1110000
CR382125_11	1100001	1210000
CR382125_12	1200001	1310000
CR382125_13	1300001	1410000
CR382125_14	1400001	1510000
CR382125_15	1500001	1610000
CR382125_16	1600001	1710000
CR382125_17	1700001	1810000
CR382125_18	1800001	1910000
CR382125_19	1900001	2010000
CR382125_20	2000001	2110000
CR382125_21	2100001	2210000
CR382125_22	2200001	2234072

Continuation (5 of 23) of CR382125 from base 400001 (CR382125 Kluyveromyces lactis strain

Query Match 90.0%; Score 18; DB 15; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19

Db 69477 AAGTAAATGCAGAAACAG 69494

RESULT 28

AL157813/C

LOCUS

DEFINITION

Human DNA sequence from clone Rp11-165D7 on chromosome 13 Contains

the CHC1L gene for chromosome condensation 1-like, 3 novel genes

and a CpG island, complete sequence.

AL157813

AL157813.11 Gi:13160187

HTG; CHC1L; CpG island.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 (bases 1 to 114519)

AUTHORS Oliver.K.

TITLE Direct Submission

JOURNAL Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk

Clone requests: clonerequests@sanger.ac.uk

On Feb 27, 2001 this sequence version replaced gi:12743755.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information

on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

chromosome 13, constructed by the Sanger Centre Chromosome 13

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr13

Rp11-165D7 is from the library RPC1-11.1 constructed by the group

of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: vegas@sanger.ac.uk

-----

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

FEATURES

source

1. 114519  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-165D7"  
/clone\_lib="RPC1-11.1"  
1  
/note="Clone left end: RP11-165D7"  
complement(join(554. .1842,7782. .7912,11212. .11351,  
13333. .13459,14315. .14505,22220. .22362,23361. .23468,  
23607. .23765,24320. .24486,26796. .26946,27177. .27332,  
33489. .33553,33899. .33994,36250. .36348,44619. .44751))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
complement(join(554. .1842,7782. .7912,11212. .11351,  
13333. .13459,14315. .14505,22220. .22362,23361. .23468,  
23607. .23765,24320. .24486,26796. .26946,27177. .27332,  
33489. .33553,33899. .33994,36250. .36348,44619. .44751))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
/product="chromosome condensation 1-like"  
/note="match: ESTs: AA116579 AF245932 AL5596966 AL702635  
AU129309 AU132093 AU138162 AV711315 AW210388 BB851525  
BF287003 BF416274 BF901505 BG027293 BI667346 BI756233  
BI918241 BI918675 BM227327 BM228676 BM681889 BM723702  
BQ184658  
match: CDNAs: AF060219 AK023010 BC029052"  
complement(554)  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
complement(577. .582)  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
complement(join(1488. .1842,3243. .3306,4031. .4164))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-003"  
complement(join(1488. .1842,3243. .3306,4031. .4164))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-003"  
/product="chromosome condensation 1-like"  
/note="match: ESTs: B1438387"  
complement(join(1702. .1842,7782. .7912,11212. .11351,  
13333. .13459,14315. .14505,22220. .22362,23361. .23468,  
23607. .23765,24320. .24486,26796. .26946,27177. .27332,  
33489. .33530))  
/gene="CHC1L"  
/locus\_tag="RP11-165D7.1-001"  
standard\_name="OTTHUMP00000018399"  
/note="match: proteins: O95199 Q99LJ7 Q9NV13"  
/codon\_start=1  
/product="chromosome condensation 1-like"  
/protein\_id="CAH70795.1"  
/db\_xref="GI:55661682"  
/db\_xref="Gene:1914"  
/db\_xref="GOA:O95199"  
/db\_xref="InterPro:IPR000210"  
/db\_xref="InterPro:IPR000408"  
/db\_xref="UniProt/TREMBL:O95199"  
/translation="MEBELPLFGSDGSGKPVQATLSSKMLDVGKWPIFSLSCEELQL  
IRQACVFGSAGNEVLYTTVNDEIFVLGTNCCGLGSDVQSTIEPRRLDSLNGKIIAC



```

LSYSGPHIVLATTEGEVFTWGNAYSQLNGTTHNLGLVPCHHISTNLSNKQVIEVACG
SVHDLVTGDEGVANGNNGSGVSTVNOPIPRRTVTCGLQNKVVVYTIACGCMCM
AYVTLGVTVGVNGQGLGNSGNQTPCVAALQGRVORVACGVYAAHTLVLTDEG
QYVANGANSYQGLGTGNKSNQSPYPTVTEKDRILEIAACHSTHTSAAKTQGGHYVM
GQCGQSVILPHLTHFSCDIDDVACFATPAVTRWLLSVDPDDHLTVAESLSKREFDNP
TADLFLVQKRYIAHKVLLKRCHEFRSSLEDNDDIVEMSEFSYVVRVAFLEYLYT
DSISLPEBAVGLLDIATFYRNLKLCQQTIKOGICEENAIALLSAARKYDAQDLE
EFCFRFCINHLTVVTQTSFAMENDHLLKNFTSKASRVGAFKN"
3512
/note="Clone_right_end: RP11-174110"
14623..14650
/note="Single clone region. Assembly confirmed by
restriction digest data."
complement(16657..18528)
/locus tag="RP11-165D7.2-001"
complement(16657..18528)
/locus tag="RP11-165D7.2-001"
/note="match: ESTs: AK058039"
complement(17675..18298)
/locus tag="RP11-165D7.2-001"
/standard_name="OTTHUMP0000018400"
/note="match: proteins: Q96DJ9"
/codon_start=1
/protein_id="CAH70794.1"
/db_xref="GI:55661681"
/db_xref="UniProt/TREMBL:O5VZX9"
/translation="MALQLRWFKHRRPFELLTAQLCVHWPHELPPDQRFGLPOHCO
PGVFLSSGPPSAPAPPALTSPRSLLGLNLAWGLSQSRLEFFLSLAACLAHHC
GOVLSSENTISHHGVKAVSHLPITSDOIRTLRWIEADATVPKPHCPPIPALAS
VPRVSTMLLSAILPSPALPCWPCACIALHTDSYSTALASPKHTWPS"
complement(join(23723..23765,24320..24486,26796..26946,
27177..27332,44619..44771))
/gene="CHC1L"
/locus tag="RP11-165D7.1-002"
complement(join(23723..23765,24320..24486,26796..26946,
27177..27332,44619..44771))
/gene="CHC1L"
/locus tag="RP11-165D7.1-002"
/product="chromosome condensation 1-like"
/note="match: ESTs: BI754136 BI754283"
complement(join(82230..82432,84699..84908))
/locus tag="RP11-165D7.3-001"
complement(join(82230..82432,84699..84908))
/locus tag="RP11-165D7.3-001"
/note="match: ESTs: AI769787"
complement(join(88565..89450,89604..89679))
/locus tag="RP11-165D7.4-001"
complement(join(88565..89450,89604..89679))
/locus tag="RP11-165D7.4-001"
/note="match: ESTs: AI307356 AI762913 AI767356 AI796352
AW469968 BM472085"
114420
/note="Clone_left_end: RP11-108P5"

ORIGIN
Query Match 90.0%; Score 18; DB 8; Length 114519;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19
|||||
30030 AAGTAAATGCAGAAACAG 30013

RESULT 29
AL833774/c
LOCUS
DEFINITION Mouse DNA sequence from clone RP23-383116 on chromosome 2, complete
sequence.
ACCESSION AL833774
VERSION AL833774.4
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)

AL833774 194045 bp DNA linear ROD 09-AUG-2002
LOCUS
DEFINITION Rattus norvegicus clone CH230-368N4, WORKING DRAFT SEQUENCE, 5
unordered pieces.
ACCESSION AC134627
VERSION AC134627.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 204712)
REFERENCE
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,

```

## ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

## Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 194045)

Tracey,A.

Direct Submission

Submitted (06-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk

On Aug 11, 2002 this sequence version replaced gi:21953137.

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: <http://www.sanger.ac.uk>

Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

-----

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP  
database can be found at  
[http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-383116 is  
from the RPCI-23 Mouse PAC Library  
constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.

## FEATURES

## Location/Qualifiers

1..194045  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/clone\_lib="RPCI-23"

## ORIGIN

Query Match 90.0%; Score 18; DB 9; Length 194045;  
Best Local Similarity 100.0%; Pred. No. 6e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACA 18

|||||

Db 94793 CAAGTAAATGCAGAAACA 94776

## RESULT 30

## AC134627

## LOCUS

## DEFINITION

Rattus norvegicus clone CH230-368N4, WORKING DRAFT SEQUENCE, 5  
unordered pieces.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## Rattus norvegicus

## Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

## Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

## Sciurognathi; Muridae; Murinae; Rattus.

## 1 (bases 1 to 204712)

## REFERENCE

## AUTHORS

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,

Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazoe, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escoto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Labow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensuhea, L., Loulaeged, H., Lozado, R., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mahoney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaoketeme, O., Okwuonu, G., Olarnpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shateman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weis, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 204712)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (28-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 204712)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On Nov 20, 2002 this sequence version replaced gi:23343609.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome

shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: KC20  
Center clone name: CH230-368N4  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 186320 bases at least Q40  
Consensus quality: 188335 bases at least Q30  
Consensus quality: 189791 bases at least Q20  
Estimated insert size: 187550; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.  
-----  
\* 1 6702: contig of 6702 bp in length  
\* 6703 6788: contig of unknown length  
\* 6789 7688: contig of 69986 bp in length  
\* 7689 7688: gap of unknown length  
\* 7689 127999: contig of 51011 bp in length  
\* 127999 127999: gap of unknown length  
\* 128000 153581: contig of 25582 bp in length  
\* 153581 153681: gap of unknown length  
\* 153681 204712: contig of 51031 bp in length.  
----- Location/Qualifiers  
1..204712  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-368N4"  
6703..6802  
/estimated\_length=unknown  
8626..9525  
/note="clone boundary  
site\_end:Spf  
end\_sequence:BZ168747"  
76789..76888  
/estimated\_length=unknown  
76889..78500  
/note="wgs\_contig"  
127900..127999  
/estimated\_length=unknown  
153582..153681  
/estimated\_length=unknown  
177120..178477  
/note="wgs\_contig"  
-----  
gap  
misc\_feature  
gap  
misc\_feature  
gap  
misc\_feature  
gap  
misc\_feature  
ORIGIN  
Query Match 90.0%; Score 18; DB 14; Length 204712;  
Best Local Similarity 100.0%; Pred. No. 6e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 3 AGTAAATGCAGAACAGG 20  
DB 134800 AGTAAATGCAGAACAGG 134817  
RESULT 31  
AC095481

AC095481	276881 bp	DNA	linear	HTG 09-MAY-2003	
Rattus norvegicus clone CH230-7K20, *** SEQUENCING IN PROGRESS ***					
5 unordered pieces.					
AC095481					
AC095481.6	GI:30467658				
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.					
Rattus norvegicus (Norway rat)					
Rattus norvegicus					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.					
1 (bases 1 to 276881)					
Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Garroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseege,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathew,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwokilemeh,O., Okwuonu,G., Olarnpusagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.					
Direct Submission					
2 (bases 1 to 276881)					
Worley, K.C.					
Direct Submission					
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
3 (bases 1 to 276881)					
Rat Genome Sequencing Consortium.					
Direct Submission					
Submitted (09-MAY-2003) Human Genome Sequencing Center, Department					

LOCUS	AC095481	276881 bp	DNA	linear	HTG 09-MAY-2003	
DEFINITION	Rattus norvegicus clone CH230-7K20, *** SEQUENCING IN PROGRESS ***					COMMENT
5 unordered pieces.						
ACCESSION	AC095481					
VERSION	AC095481.6	GI:30467658				
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.					
SOURCE	Rattus norvegicus (Norway rat)					
ORGANISM	Rattus norvegicus					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Rattus.					
AUTHORS	Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Garroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseege,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mathew,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwokilemeh,O., Okwuonu,G., Olarnpusagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.					
TITLE	Direct Submission					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 276881)					
AUTHORS	Worley, K.C.					
TITLE	Direct Submission					
JOURNAL	Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
REFERENCE	3 (bases 1 to 276881)					
AUTHORS	Rat Genome Sequencing Consortium.					
TITLE	Direct Submission					
JOURNAL	Submitted (09-MAY-2003) Human Genome Sequencing Center, Department					

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
On May 9, 2003 this sequence version replaced gi:24941064.	
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.	
----- Genome Center of Medicine Center: Baylor College of Medicine Center code: BCM Web site: http://www.hgsc.bcm.tmc.edu/ Contact: hgsc-help@bcm.tmc.edu ----- Project Information Center project name: GCKW Center clone name: CH230-7K20 ----- Summary Statistics Assembly program: Atlas; Consensus quality: 22954 bases at least Q40 Consensus quality: 228004 bases at least Q30 Consensus quality: 230763 bases at least Q20 Estimated insert size: 233155; sum-of-contigs estimation Quality coverage: 6x in Q20 bases; sum-of-contigs estimation -----	
* NOTE: Estimated insert size may differ from sequence length * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html) * NOTE: This is a 'working draft' sequence. It currently * consists of 5 contigs. The true order of the pieces * is not known and their order in this sequence record is * arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence. * as soon as it is available and the accession number will * be preserved.	
* 1 241308: contig of 241308 bp in length * 241309 241408: gap of unknown length * 241409 270462: contig of 29054 bp in length * 270463 270562: gap of unknown length * 270563 271681: contig of 1119 bp in length * 271682 271781: gap of unknown length * 271782 272810: contig of 1029 bp in length * 272811 272910: gap of unknown length * 272911 276881: contig of 3971 bp in length.	
FEATURES source 1. .276881 /organism="Rattus norvegicus" /mol_type="genomic DNA" /db_xref="taxon:10116" /clone="CH230-7K20" misc_feature 1. .1235 /note="wgs contig" misc_feature 63196. .65045 /note="wgs contig" misc_feature 167109. .168091 /note="wgs contig" gap 241309. .241408 /estimated_length=unknown gap 270463. .270562 /estimated_length=unknown gap 271682. .271781 /estimated_length=unknown gap 272811. .272910 /estimated_length=unknown	
ORIGIN	
Query Match	90.0%; Score 18; DB 14; Length 276881;

of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA		Center: Genome Center	
On May 9, 2003 this sequence version replaced gi:24941064.		Center code: BCM	
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.		Web site: http://www.hgsc.bcm.tmc.edu/	
		Contact: hgsc-help@bcm.tmc.edu	
		----- Project Information	
		Center project name: GCKW	
		Center clone name: CH230-7K20	
		----- Summary Statistics	
		Assembly program: Atlas	
		Consensus quality: 222954 bases at least Q40	
		Consensus quality: 228004 bases at least Q30	
		Consensus quality: 230763 bases at least Q20	
		Estimated insert size: 233155; sum-of-contigs estimation	
		Quality coverage: 6x in Q20 bases; sum-of-contigs estimation	
		-----	
		* NOTE: Estimated insert size may differ from sequence length (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).	
		* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.	
		* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
		* 1 241308: contig of 241308 bp in length	
		* 241309 241408: gap of unknown length	
		* 241409 270462: contig of 29054 bp in length	
		* 270463 270562: gap of unknown length	
		* 270563 271681: contig of 1119 bp in length	
		* 271682 271781: gap of unknown length	
		* 271782 272810: contig of 1029 bp in length	
		* 272811 272910: gap of unknown length	
		* 272911 276881: contig of 3971 bp in length.	
		----- Location/Qualifiers	
		1. .276881	
		/organism="Rattus norvegicus"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:10116"	
		/clone="CH230-7K20"	
		1. .1235	
		/note="wgs contig"	
		63196. .65045	
		/note="wgs contig"	
		167109. .168091	
		/note="wgs contig"	
		241309. 241408	
		/estimated_length=unknown	
		270463. .270562	
		/estimated_length=unknown	
		271682. .271781	
		/estimated_length=unknown	
		272811. 272910	
		/estimated_length=unknown	
		-----	
		90.0%; Score 18; DB 14; Length 276881;	

Best Local Similarity 100.0%; Pred. No. 5.9e+02; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGTAATAATGCAGAAACAGG 20  
|||||  
Db 247941 AGTAATAATGCAGAAACAGG 247958

RESULT 32  
G52355/c  
LOCUS SHGC-85070 Human Homo sapiens STS genomic, sequence tagged site.  
DEFINITION G52355  
ACCESSION G52355  
VERSION G52355.1 GI:5223682  
KEYWORDS STS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE  
AUTHORS Olivier, M. and Cox, D.R.  
TITLE Unpublished, Olivier, M., Cox, D.R. (2000)  
JOURNAL Unpublished (2000)  
COMMENT

Contact: Michael Olivier, David R. Cox  
Stanford Human Genome Center  
Stanford University School of Medicine  
2005 Miranda Ave. 2nd Fl., Palo Alto, CA 94025, USA  
Tel: (650) 320-5800  
Fax: (650) 320-5801  
Email: olivier@shgc.stanford.edu  
Primer A: GGAGAACTCGGTGAGAGAAATT  
Primer B: GTGCTAAATGACACCCACATGAA  
STS size: 286  
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
AmpliTaq Gold Polymerase: 0.07 units/ul  
Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

BAC ends sequenced at TIGR from the RPC111 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES  
source

1. .491  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/map="12"  
/clone\_lib="Human"  
191. .476  
191. .213  
complement(454. .476)

STS  
primer\_bind  
primer\_bind

Query Match 87.0%; Score 17.4; DB 10; Length 491;  
Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACAG 19  
|||||  
Db 242 CAAGTAATGCAGAAACAG 224

RESULT 33  
G90172/c  
LOCUS S210P603RB3.T0 BALB/cByJ Mus musculus STS genomic, sequence tagged site.  
DEFINITION G90172  
ACCESSION G90172  
VERSION G90172.1 GI:22740928  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS Wade, C.  
TITLE Polymorphism Structure in the Mouse  
JOURNAL Unpublished (2002)  
COMMENT

Contact: Kerstin Lindblad-Toh  
Whitehead Institute for Biomedical Research, Center for Genome Research  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172521477  
Fax: 6172580903  
Email: kersli@genome.wi.mit.edu  
Primer A: None  
Primer B: None  
STS size: 549  
Protocol:

WGS-discovery: Paired-end low-coverage whole genome shotgun reads were generated from 129S1/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP detection was carried out by SSAHA-SNP. 225,000 reads were annotated as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%.

FEATURES  
source

1.549  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="BALB/cByJ"  
/db\_xref="taxon:10090"  
/map="- 1 22-488 157394789-157395256"  
/clone\_lib="BALB/cByJ"  
<1. .549

STS  
ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 549;  
Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAATGCAGAAACAGG 20  
|||||  
Db 322 AAGTAATGCAGAAACAGG 304

RESULT 34  
BV558070  
LOCUS gmm55f08.g1 Clint Pan troglodytes versus STS genomic, sequence tagged site.  
DEFINITION BV558070  
ACCESSION BV558070.1 GI:62449091  
VERSION BV558070  
KEYWORDS STS.  
SOURCE Pan troglodytes versus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

LOCUS DEFINITION	BV586373 G591P631877RH10.T0 Clint Pan troglodytes versus STS genomic, sequence tagged site.	737 bp	DNA	linear	STS 12-APR-2005
ACCESSION	BV586373				
VERSION	BV586373.1	GI:62502102			
KEYWORDS	STS.				
SOURCE	Pan troglodytes verus				
ORGANISM	Pan troglodytes verus Pan troglodytes verus Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pan. 1 (bases 1 to 737) Mikkelsen,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and Jaffe,D.B. Initial Sequence of the Chimpanzee Genome and Comparison with the Human Genome Unpublished (2005)				
REFERENCE AUTHORS	Contact: Michael C. Zody Broad Institute of MIT and Harvard 320 Charles Street, Cambridge, MA 02141, USA Tel: 6172580933 Fax: 6172580903 Email: mczody@broad.mit.edu Primer A: No sequence submitted Primer B: No sequence submitted STS size: 737				
TITLE	Protocol: 23,021,928 chimpanzee whole genome shotgun reads were aligned to the Human genome NCBI Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors, including Clint (Pan troglodytes verus), 3 other Pan troglodytes verus chimps (Donald,Karlien,Yvonne), 3 Pan troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps of unknown origin (Gon,Unknown Chimp). Common names: Pan troglodytes verus is the western chimp and Pan troglodytes troglodytes is the central chimp. To be included in chimpanzee SNP discovery, a read must be at least 500bp in length, at least 50% of its base calls must have Phred score >= 20, at least 30% of its base calls must satisfy SNQS(30,25)(single strand NQS, the base in question has Phred score >= 30, the surrounding 10 bases in the read have Phred score >= 25), and the read must have at least 200 bp SNQS(30,25) bases. Reads not uniquely placed in the genome and read pairs whose two ends were not consistently placed were discarded. After above filtering, NQS(30,25) standard was applied to all pairs of overlapping reads to call NQS bases and SNPs. Alignments (between two reads) with less than 100 NQS bases or with SNP rate > 0.01 were discarded. To exclude alignment between two copies of a single read, comparisons between two reads that share 95% of their genome alignments (>=95% bases of read A and >=95% bases of read B were placed at the same locus of human genome) were discarded.				
JOURNAL COMMENT	Location/Qualifiers 1..737 /organism="Pan troglodytes verus" /mol_type="genomic DNA" /sub_species="verus" /db_xref="taxon:37012" /clone_lib="Clint" <1..>737				
FEATURES	source				
ORIGIN	Query Match 87.0%; Score 17.4; DB 10; Length 723; Best Local Similarity 94.7%; Pred. No. 1.5e+03; Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0; Qy 2 AAGTAAATGCAGAACAGG 20       Db 641 AAGTAAATGCAGAACAGG 659 				
STS					
ORIGIN	Query Match 87.0%; Score 17.4; DB 10; Length 737; Best Local Similarity 94.7%; Pred. No. 1.5e+03; Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0; Qy 2 AAGTAAATGCAGAACAGG 20       Db 641 AAGTAAATGCAGAACAGG 659 				
RESULT 35					
BV586373/c					

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAACAGG 20  
 |||||  
 Db 367 AAGTAAATGCAGAACAGG 349

RESULT 36  
 BV576492  
 LOCUS  
 DEFINITION BV576492 800 bp DNA linear STS 12-APR-2005  
 G591P608338RB5.TO Clint Pan troglodytes verus STS genomic, sequence  
 tagged site.

ACCESSION BV576492  
 VERSION BV576492.1 GI:62492221  
 KEYWORDS STS.  
 SOURCE Pan troglodytes verus

ORGANISM Pan troglodytes verus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Pan. (bases 1 to 800)

REFERENCE 1 (bases 1 to 800)  
 AUTHORS Mikkelson,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and  
 Jaffe,D.B.

TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the  
 Human Genome

JOURNAL Unpublished (2005)

COMMENT Contact: Michael C. Zody  
 Broad Institute of MIT and Harvard  
 320 Charles Street, Cambridge, MA 02141, USA  
 Tel: 6172580933  
 Fax: 6172580930  
 Email: mczody@broad.mit.edu  
 Primer A: No sequence submitted  
 Primer B: No sequence submitted  
 STS size: 800

Protocol:  
 23,021,928 chimpanzee whole genome shotgun reads were aligned to  
 the Human genome NCBI  
 Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,  
 including Clint (Pan

troglodytes verus), 3 other Pan troglodytes verus chimps  
 (Donald,Karlén,Yvonne), 3 Pan

troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps  
 of unknown origin  
 (Gon,Unknown Chimp). Common names: Pan troglodytes verus is the  
 western chimp and Pan

troglodytes troglodytes is the central chimp. To be included in  
 chimpanzee SNP discovery, a  
 read must be at least 500bp in length, at least 50% of its base  
 calls must have Phred

score >= 20, at least 30% of its base calls must satisfy  
 SNQS(30,25)(single strand NQS, the  
 base in question has Phred score >= 30, the surrounding 10 bases in  
 the read have Phred

score >= 25), and the read must have at least 200 bp SNQS(30,25)  
 bases. Reads not uniquely  
 placed in the genome and read pairs whose two ends were not  
 consistently placed were  
 discarded. After above filtering, NQS(30,25) standard was applied  
 to all pairs of

overlapping reads to call NQS bases and SNPs. Alignments (between  
 two reads) with less  
 than 100 NQS bases or with SNP rate > 0.01 were discarded. To  
 exclude alignment between two  
 copies of a single read, comparisons between two reads that share  
 95% of their genome

alignments (>=95% bases of read A and >=95% bases of read B were  
 placed at the same locus  
 of human genome) were discarded.  
 Location/Qualifiers  
 1. .800  
 /organism="Pan troglodytes verus"

FEATURES  
 source

/mol\_type="genomic DNA"  
 /sub\_species="verus"  
 /db\_xref="taxon:37012"  
 /clone\_lib="Clint"  
 <1..>800

Query Match 87.0%; Score 17.4; DB 10; Length 800;  
 Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAACAGG 20  
 |||||  
 Db 698 AAGTAAATGCAGAACAGG 716

RESULT 37  
 BV628207  
 LOCUS  
 DEFINITION BV628207 800 bp DNA linear STS 15-APR-2005  
 S217962301FA10.TO Noemie Pan troglodytes troglodytes STS genomic,  
 sequence tagged site.

ACCESSION BV628207  
 VERSION BV628207.1 GI:62618385  
 KEYWORDS STS.  
 SOURCE Pan troglodytes troglodytes

ORGANISM Pan troglodytes troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homidae; Pan. (bases 1 to 800)

REFERENCE 1 (bases 1 to 800)  
 AUTHORS Mikkelson,T.S., Hillier,W.L., Eichler,E.E., Zody,M.C. and  
 Jaffe,D.B.

TITLE Initial Sequence of the Chimpanzee Genome and Comparison with the  
 Human Genome

JOURNAL Unpublished (2005)

COMMENT Contact: Michael C. Zody  
 Broad Institute of MIT and Harvard  
 320 Charles Street, Cambridge, MA 02141, USA  
 Tel: 6172580933  
 Fax: 6172580930  
 Email: mczody@broad.mit.edu  
 Primer A: No sequence submitted  
 Primer B: No sequence submitted  
 STS size: 800

Protocol:  
 23,021,928 chimpanzee whole genome shotgun reads were aligned to  
 the Human genome NCBI  
 Build 34 (hg16,July 2003). Chimp WGS reads were from 9 donors,  
 including Clint (Pan

troglodytes verus), 3 other Pan troglodytes verus chimps  
 (Donald,Karlén,Yvonne), 3 Pan

troglodytes troglodytes chimps (Noemie,Masuku,Clara) and 2 chimps  
 of unknown origin  
 (Gon,Unknown Chimp). Common names: Pan troglodytes verus is the  
 western chimp and Pan

troglodytes troglodytes is the central chimp. To be included in  
 chimpanzee SNP discovery, a  
 read must be at least 500bp in length, at least 50% of its base  
 calls must have Phred

score >= 20, at least 30% of its base calls must satisfy  
 SNQS(30,25)(single strand NQS, the  
 base in question has Phred score >= 30, the surrounding 10 bases in  
 the read have Phred

score >= 25), and the read must have at least 200 bp SNQS(30,25)  
 bases. Reads not uniquely  
 placed in the genome and read pairs whose two ends were not  
 consistently placed were  
 discarded. After above filtering, NQS(30,25) standard was applied  
 to all pairs of

overlapping reads to call NQS bases and SNPs. Alignments (between  
 two reads) with less  
 than 100 NQS bases or with SNP rate > 0.01 were discarded. To  
 exclude alignment between two  
 copies of a single read, comparisons between two reads that share  
 95% of their genome

alignments (>=95% bases of read A and >=95% bases of read B were  
 placed at the same locus  
 of human genome) were discarded.  
 Location/Qualifiers  
 1. .800  
 /organism="Pan troglodytes verus"

FEATURES  
 source

exclude alignment between two  
copies of a single read, comparisons between two reads that share  
95% of their genome  
alignments (>=95% bases of read A and >=95% bases of read B were  
placed at the same locus  
of human genome) were discarded.

## FEATURES

source  
1. .800  
Location/Qualifiers  
/organism="Pan troglodytes troglodytes"  
/mol\_type="genomic DNA"  
/sub\_species="troglodytes"  
/db\_xref="taxon:37011"  
/clone\_lib="Noemie"  
<1. .>800

## STS

## ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 800;  
Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAATGCAGAAACAGG 20

DB 412 AAGTGAATGCAGAAACAGG 430

## RESULT 38

## BV594335/c

## LOCUS

DEFINITION BV594335 806 bp DNA linear STS 12-APR-2005  
taw08f01.b1 Clint Pan troglodytes versus STS genomic, sequence  
tagged site.

## ACCESSION

## BV594335

## VERSION

KEYWORDS BV594335.1 GI:62510064

## SOURCE

## ORGANISM

Pan troglodytes versus  
Pan troglodytes versus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eumetazoa; Euarchontoglires; Primates; Catarrhini;  
Hominoidea; Pan.

## REFERENCE

## AUTHORS

## TITLE

Initial Sequence of the Chimpanzee Genome and Comparison with the  
Human Genome  
Unpublished (2005)

## JOURNAL

## COMMENT

Contact: Michael C. Zody  
Broad Institute of MIT and Harvard  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172580933  
Fax: 6172580903  
Email: mczody@broad.mit.edu  
Primer A: No sequence submitted  
Primer B: No sequence submitted  
STS size: 806  
Protocol:

23,021,928 chimpanzee whole genome shotgun reads were aligned to  
the Human genome NCBI  
Build 34 (hg16, July 2003). Chimp WGS reads were from 9 donors,  
including Clint (Pan  
troglodytes versus), 3 other Pan troglodytes versus chimps  
(Donald, Karlén, Yvonne), 3 Pan  
troglodytes troglodytes chimps  
(Noemie, Masuku, Clara) and 2 chimps  
of unknown origin  
(Gon, Unknown Chimp). Common names: Pan troglodytes versus is the  
western chimp and Pan  
troglodytes troglodytes is the central chimp. To be included in  
chimpanzee SNP discovery, a  
read must be at least 500bp in length, at least 50% of its base  
calls must have Phred  
score >= 20, at least 30% of its base calls must satisfy  
SNOS(30,25)(single strand NQS, the  
base in question has Phred score >= 30, the surrounding 10 bases in  
the read have Phred

score >= 25), and the read must have at least 200 bp SNQS(30,25)  
bases. Reads not uniquely  
placed in the genome and read pairs whose two ends were not  
consistently placed were  
discarded. After above filtering, NQS(30,25) standard was applied  
to all pairs of  
overlapping reads to call NQS bases and SNPs. Alignments (between  
two reads) with less  
than 100 NQS bases or with SNP rate > 0.01 were discarded. To  
exclude alignment between two  
copies of a single read, comparisons between two reads that share  
95% of their genome  
alignments (>=95% bases of read A and >=95% bases of read B were  
placed at the same locus  
of human genome) were discarded.

## FEATURES

## source

1. .806  
Location/Qualifiers  
/organism="Pan troglodytes versus"  
/mol\_type="genomic DNA"  
/sub\_species="versus"  
/db\_xref="taxon:37012"  
/clone\_lib="Clint"  
<1. .>806

## STS

## ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 806;  
Best Local Similarity 94.7%; Pred. No. 1.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAATGCAGAAACAGG 20

DB 549 AAGTAATGCAGAAACAGG 531

## RESULT 39

## BC063923/c

## LOCUS

## DEFINITION

BC063923 2081 bp mRNA linear VRT 23-AUG-2004  
Xenopus tropicalis hypothetical protein MGC76215, mRNA (cDNA clone  
MGC:76215 IMAGE:5336233), complete cds.

## ACCESSION

## BC063923

## VERSION

## BC063923.1

## KEYWORDS

## MGC.

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

Xenopus tropicalis (Silurana tropicalis)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 2081)  
Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,  
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Donaldson, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raja, S.S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Bouffard, G., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Santacruz, A., Blakeley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalios, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
JOURNAL  
PUBMED  
2 (bases 1 to 2081)  
REFERENCE  
Klein, S. and Gerhard, D.S.  
Direct Submission  
TITLE



JOURNAL	Submitted (08-DEC-2003) National Institutes of Health, Xenopus Gene Collection (XGC), National Institute of Child Health and Human Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD 20892-7510, USA			REFERENCE	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Homnidae; Homo.		
REMARK	NIH-MGC Project			AUTHORS	Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayaashi, K., Sato, H., Nagai, K., Kimura, K., Makita, H., Sekine, M., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Iehi, S., Yamamoto, J., Saito, K., Kawai, Y., Isono, Y., Nakamura, Y., Negahari, K., Murakami, K., Yasuda, T., Iwayanagi, T., Wagatsuma, M., Shiratori, A., Sudo, H., Hoshino, T., Kaku, Y., Kodaira, H., Kondo, H., Sugawara, M., Takahashi, M., Kanda, K., Kamiyama, K., Katsuta, S., Sato, K., Kikkawa, E., Omura, Y., Abe, K., Kamihara, K., Ishibashi, T., Yamashita, H., Tanikawa, M., Yamazaki, M., Ninomiya, K., Ishibashi, T., Yamashita, H., Murakawa, K., Fujimori, K., Tanai, H., Kimata, M., Watanabe, M., Hirakawa, S., Chiba, Y., Ishida, S., Ono, Y., Takiguchi, S., Watanabe, S., Yoshida, M., Hoteuta, T., Kusano, J., Kanehori, K., Takahashi, Fujii, A., Hara, H., Tanase, T., Nomura, Y., Togii, S., Komai, F., Hara, R., Takouchi, K., Arita, M., Imose, N., Mueashino, K., Yuuki, H., Oshima, A., Sasaki, N., Aotsuka, S., Yoshikawa, Y., Matsunawa, H., Ichihara, T., Shiohata, N., Sano, S., Moriya, S., Momiyama, H., Satoh, N., Takami, S., Terashima, Y., Suzuki, O., Nakagawa, S., Senoh, A., Mizoguchi, H., Goto, Y., Shimizu, F., Wakebe, H., Hishigaki, H., Watanabe, T., Sugiyama, A., Takemoto, M., Kawakami, B., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Fujimori, Y., Komiyama, M., Tashiro, H., Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M., Ohmori, Y., Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, F., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Sasaki, M., Togaishi, T., Oyama, M., Hata, H., Watanabe, M., Komatsu, T., Mizushima-Sugano, J., Satoh, T., Shirai, Y., Takahashi, Y., Nakagawa, K., Okumura, K., Nagase, T., Nomura, N., Kikuchi, H., Masuho, Y., Yamashita, R., Nakai, K., Yada, T., Nakamura, Y., Ohara, O., Isogai, T. and Sugano, S.		
FEATURES	Location/Qualifiers			TITLE	Complete sequencing and characterization of 21,243 full-length human cDNAs		
source	1..2081			JOURNAL	Nat. Genet. 36 (1), 40-45 (2004)		
gene	/organism="Xenopus tropicalis"			PUBMED	14702039		
CDS	/mol_type="mRNA"			REFERENCE	Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Iehi, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Negahari, K., Masuho, Y., Nagai, K. and Isogai, T.		
ORIGIN	/db_xref="GeneID:394910"			AUTHORS	Isogai, T. and Yamamoto, J.		
Query Match	87.0%; Score 17.4; DB 5; Length 2081;			JOURNAL	Direct Submission		
Best Local Similarity	94.7%; Pred. No. 1.4e+03;			REFERENCE	Submitted (15-JUL-2003) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan		
Matches	18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			AUTHORS	(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)		
Qy	1 CAACTAATGCAGAACAG 19			JOURNAL	NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center; National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.		
Db	409 CAACTAATGCAGAACAG 391			COMMENT	NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center; National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.		
RESULT 40	AKI23724/c			FEATURES	Location/Qualifiers		
LOCUS	AKI23724			source	1..2286		
DEFINITION	Homo sapiens cDNA FLJ41730 fis, clone HLUNG2017262.			organism="Homo sapiens"	/organism="Homo sapiens"		
ACCESSION	AKI23724			/mol_type="mRNA"	/mol_type="mRNA"		
VERSION	AKI23724.1			/db_xref="taxon:9606"	/db_xref="taxon:9606"		
KEYWORDS	oligo capping; fis (full insert sequence).			/clone="HLUNG2017262"	/clone="HLUNG2017262"		
SOURCE	Homo sapiens (human)			/tissue_type="lung"	/tissue_type="lung"		
ORGANISM	Homo sapiens			/clone_lib="HLUNG2"	/clone_lib="HLUNG2"		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				/note="cloning vector: pME18SFL3"	/note="cloning vector: pME18SFL3"		
Query Match	87.0%; Score 17.4; DB 8; Length 2286;			ORIGIN	Query Match		
Best Local Similarity	94.7%; Pred. No. 1.4e+03;				87.0%; Score 17.4; DB 8; Length 2286;		
Matches	18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;						









REFERENCE AUTHORS TITLE JOURNAL	Hominidae: Homo. 1 (bases 1 to 37170) Hunt, S. Direct Submission Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vega@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C.elegans/wormpep/LL22NC01-129H9 is from the human chromosome 22-specific cosmid library LL22NC01, constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550 under the auspices of the National Laboratory Gene Library Project sponsored by the US Department of Energy. The source of the flow sorted chromosomes was a human/hamster hybrid containing chromosomes Y, 22 and 9. VECTOR: lawrie16 This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr22 ----- Genome Center Center: Wellcome Trust Sanger Institute Center code: SC Web site: http://www.sanger.ac.uk Contact: vega@sanger.ac.uk -----
COMMENT	This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. Location/Qualifiers 1. .37170 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="22" /clone="LL22NC01-129H9" /clone_lib="LL22NC01"
FEATURES source	Query Match 87.0%; Score 17.4; DB 8; Length 37170; Best Local Similarity 94.7%; Pred. No. 1.3e+03; Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  QY 2 AAGTAAATGCAGAACAGG 20       Db 12392 AAGTAAATGCAGAACAGG 12374  RESULT 49 AC079678/c LOCUS DEFINITION Arabidopsis thaliana chromosome 1 BAC F9E11 genomic sequence, complete sequence. AC079678 VERSION AC079678.1 GI:9989658 KEYWORDS HTG. SOURCE Arabidopsis thaliana (thale cress) ORGANISM Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 39601) Lin.X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J., Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M. Arabidopsis thaliana chromosome 1 BAC F9E11 genomic sequence
ORIGIN	Unpublished 2 (bases 1 to 39601) Town, C.D. and Kaul, S. Direct Submission Submitted (07-SEP-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@igr.org 3 (bases 1 to 39601) Town, C.D. and Kaul, S. Direct Submission Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@igr.org Address all correspondence to: atetigr.org  BAC clone F9E11 is from Arabidopsis thaliana chromosome 1 clone. Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky, http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, http://www.tigr.org/softlab/glimmer/htm/glimmerm.html, and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/cgi.shtml). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Location/Qualifiers 1. .39601 /organism="Arabidopsis thaliana" /mol_type="genomic DNA" /db_xref="taxon:3702" /chromosome="1" /clone="F9E11" /ecotype="Columbia" 1096. .4664 /gene="F9E11.4" /note="similar to GI:1184077 from [Lycopersicon pimpinellifolium]" join(<1096.1465,1789.1944,2097.2171,2318.2431,2482.>.4664) /gene="F9E11.4" join(1096.1465,1789.1944,2097.2171,2318.2431,2482.4664) /gene="F9E11.4" /codon_start=1 /product="disease resistance protein, putative; 1096-4664" /protein_id="AAG51871.1" /db_xref="GI:12323812" /translation="MEGVFLGHNLIVWLMGLQHGKYSCIDBEKIALFELRKHMS RTESVLPWTNTDTSDCCRWKGVACNRVGRVTEISFGGLSLKSLNLSLHHP EDVRSNLSSRCSGLPDVGEYKSLRKLRKLEILDASKNFNNSIPHFLSAATSLTT LFRSNNMDGSPAKELRDLTNLSELLDLSNRNFGSIPQIELSRUKALKALDSNEF SGSMELQGFCTDLFFISQGI CELNNMQELDLSQNKLVGHPCLSLTSGURVLDLS SNKLTGTVPSSLSGLQSLYLSLFDNDFEGSFSGSLANLSNLMVLKLCSSKSSQLVL SESSWKPKQLSVIALRSCNKKVPHLLHQDLHVDLSNNIWSGLKPSWLLANNTK LKVLQNLNFTSQIPKSAHNLFLDVANDFNHLPENIGWIPHLRYLNTSKNPF QENLPSSLGNGNGIQYMDLSNSFHNLPBSFVNGCYSMALLKSHNKLSELRPEST NFTNLGDFMNNLFTKIGQGLASLNLLELLDNNNLTGVIPIWSWIGELPSLTALLI SDNTLKGDI PMSLNFSLQLLDLSANSLSGVIPQHDNRGVLLQLQNKLSGTIPD TLLANRILDLNRNRFSGIKPEF INIQNISILLRGNNTQGI PHQLCGLSNIQLLID SNRNLRTIPCLSNSTSGFGKCTSYDYDFGISFPSPDVFNGSLHSDPNSKNGGIY FKSLLTLDPLSMYKAATQTKIBFATKHYDAYMGNLKLLFGMDLSNELSGEIPVE FGGLLELRALNLSHNNLSGVTPKSISSMEKMFSLSPNLQGRIPSLTBLTSLSVF
FEATURES source	gene  mRNA  CDS

gene  
KVNHNLSGVIPOGRQNTFDAESYFGNRLLCQPTNRSNNNSYEADNGVEADESI  
IDMVSFVLFGAFAAYVILIGILASLSDSPWRFYKVDAFIKVRNLLL"  
/gene="F9E11.6"  
/notes="similar to GI:1184077 from [Lycopodium  
pimpinellifolium]"  
join(<6346..6733,7017..7172,7273..7341,7661..7735,  
7890..>10057)  
/gene="F9E11.6"  
/join(6346..6733,7017..7172,7273..7341,7661..7735,  
7890..10057)  
/codon\_start=1  
/product="disease resistance protein, putative;  
6346-10057"  
/protein\_id="AAG51872.1"  
/db\_xref="GI:12323813"  
/translation="MERKPSGQNLIVMVLVLRGKYKCCIEKERRKALLEKKYMS  
KPADWGLDSVLPWTNDTKSNCCRWGLKCNQTSGRRIEISIGQTNFKESLLNLSLL  
HPEELRSNLNGEINYEENGLFDDVEGESLRRLRNLEITLDSNSFNNSIPFPLNA  
ATSLTLTQSNVIGGLPIKELKNLTKLELDRSRGNGSTPELKVLNLEVLGLA  
WNHLDGPIKEVCEMKNLRQLDLRGNYFEGQLPVCLGNLKNLRVLDSNQLSGLP  
ASNSLESLEYSLDSNNEFFESLNLPLANTLTKVFRLSSTSEMLOVETESNWLKPF  
QLTVAALPFCGLKIPNLFVYQTNRLVDLSNRLSGDIPTWLLENNPELKVLSMG  
SFTIPOITVHKLQVLDIFSANDITGVLDNIGVLRPLIHMGNSHNGQNLPSMG  
EMNDISPLDLYNNFSGELPRSLTGCFSLITLQSHNSPSPILPIQTRLTSLVLR  
MNNDITGEGVGLRVLNLSIFDASNNRLTGLISSIPDSSHLMILLNNLLEGT  
LPPSLIAHLNLDLSNGLSGLDPSVVNSMYGKIFLHNSFTGCPPLVTILENAV  
ILDRNNKSGSTPQVNTGKMITLLRGNLNGTSIPRKLCDUTSIRLLDSDNKUNG  
VLPCLNHLSTELGEGIGSGFQSEISFGDSLQMEFYRSTFLVDEFMYYDSYMLF  
IEFAAKORYDSFGSDLDYNNGLDLSNLSGVPYPAELGDLKRLAINLNRLLSSSI  
PANFKLKDJESLSDLYNNQGNIPHQLTNLTSLAVFNVSFNNLSGIIPOGGQNTFN  
DNLYGNPLLCGTPDRSCGKNKTEADNGGSEEDDDDEAADIWVLYWTGTSTY  
AIALIGLVLMCFDCPWRRTWLCIVDAFIASGSMFS"  
11609..15699  
/gene="F9E11.7"  
/notes="similar to GI:1184077 from [Lycopodium  
pimpinellifolium]"  
join(<11609..12002,12342..12497,12658..12732,12853..13008,  
13095..13169,13288..13392,13454..>15699)  
/gene="F9E11.7"  
/join(11609..12002,12342..12497,12658..12732,12853..13008,  
13095..13169,13288..13392,13454..15699)  
/gene="F9E11.7"  
/codon\_start=1  
/product="disease resistance protein, putative;  
11609-15699"  
/protein\_id="AAG51873.1"  
/db\_xref="GI:12323814"  
/translation="MEGKPLGLQYLICVILLGLQHGKYKCCIEKERRKALLEKAFILP  
LNAGEMNDVLSNTDTSKCCQMGVCEKGRITNIAFGIFGIENPLNLSLLH  
PFEDVRLSDLSRRSCDCGFGFLDDVEGKSLSRURNLEILDLSHRRFNNSIFPLP  
NAATSLTLTFLTYNNMHSPLVKEFKDLTNLEHLDLRGNFNGSIPTQDYNLSRRFRK  
LEILDNDLNFNSRIFPFLNSATSLKSLAGNMGPPPAKELRDLTNTVNLDLDSRN  
RNGSIIPVALFALRKALDLDSDNEFSSVVELQGFKAFTKPLSGTCKPWNMEGKSL  
NNKLAGOPLCLTSLGRLVLDLSNQLTGNVSALANLESLEVLSPGKNFEGFSL  
GLLANLSKLRDLSQNSLEVEFETSWRKPQLVIALRSCLNLEKVPFLHQLKDL  
HVDLSNDTHGNFSPWLENNKTLKVLLONNSTFTSFLPKSAHNLFLNVNKNF  
HLFLQNFQWTLPHVCLVYNGFQNLPSLDNKKSIIEFLDLSHNRPHGKLPFRFLK  
GCYNLTILKLSHNLKSGVPEAAANFTLRWMSMDNLLFTGNIGKFRSLPSNLVDI  
SNKKLTGVPISWICERGLPALQLSNNMGEIPTSLEFNISYQLQLDLSNRLSGDIP  
PHVSIYHGAVLLQNNLSGVIPTDILLNVLDLRNRLSGNLPEFINTONISILL  
LRGNFTGQIPQPCUSNQLLOLNNKNGSIPLCSLNTSFGRLKGDSDSYRDVPS  
RFGTAKDPVTFESLMDIDEFNMVNETNSQTKIEPATKRYDAYDMGNLKLFGMDLSE  
NELSGEPVLEGLGLVEALNLSHNLGVSILEFSGKLVESLDLSLFPNLQGGPIQ  
LTDMSLAVFNVSNNLSGVIPOGRQNTFDQSYFGNRLLCQPTNRSNNNSYEADNGVEADESI  
IDMVSFVLFGAFAAYVILIGILASLSDSPWRFYKVDAFIKVRNLLL"  
/gene="F9E11.1"  
/complement(join(<16044..19955))  
/gene="F9E11.1"

gene  
complement(join(16103..16959,17177..17312,17406..19292,  
19382..19452,19955..20090))  
/gene="F9E11.1"  
/codon\_start=1  
/product="unknown protein; 20090-16103"  
/protein\_id="AAG51874.1"  
/db\_xref="GI:12323815"  
/translation="MAAKLLHSLADSDADLQKQIGCMNGIQFIPDRHHVLTGRKSLT  
LGNGAININIERSDVTIYQKQETDSNTGGNVKRRVSTESSRVSFSSSSSSPSS  
SFNGQVQDASAYDRANQESPTSDPEMTEGNGFSLGLDLDRDVRDVRMREARGLL  
SPTMTREEVQRQSRREDSPRYGLKQSTPMDLNFESFVLARLRETSHQYNELGMKDA  
PRYSVDHDTLSKROKLELRLSLDSRERATNSVDPSKLSLSEFSSSCSSSKK  
RPPSVAKUMLETTLPGLGRDIIHQFGLNKTNISQNDPPRSRLREKLNRAIRFS  
PSPRSGLKDPASPRWRNSDFVMKLSNTRFVPEPAPWKHADNRNVLQKQASMPVKAK  
PYEAPNFPPTVYSSEMERLNDLEFGHSGKDLRALQILSEMSQSGFLDTEKQOQSTNF  
AVQDYRENSATSNHMSRTRVSQSSSNQVQSPIVAKPAKLVKAGIPASSLIP  
IHSLTGKIKIRREKDDKCTASNSKRVTKDCSPGNRAECSSTSSFDKKSQSRNRS  
SKPQVQVSEKASKSGSVSPRLQOKLEYDKSRPPTPDSSKSRKPSNQQLVEST  
PGRRRPKQKSLQVDDQLSQNSNRSTSSHGICQSTSEASACVEKSTEADGGKSP  
SVTEAAKAVSNLMQNKSPFSESDGLSANLSLVALEHSPISVLDASTIRETSPV  
KTQADHIVGNVAHDFGCENADPTTDHRYTSEILLASGLLLRDIGSGLTTLQHPSGHP  
LNSSHDEAQDYIASLCENADPTTDHRYTSEILLASGLLLRDIGSGLTTLQHPSGHP  
INPELFFVLEQTKGSGSTHLLHKEESKVLKNEKLNKLVDFLVNLEILVEKLSVEAT  
NLMKSYAKVTKVASAQLLKEKLSAITEITQKQAKRSENFLESEDDDLKSTLAD  
VTIRSGNADFGSEMSGLVDLVERLVKOLVNEIVHETSRLQAKSGRRRLTADQ"  
complement(23758..26726)  
/gene="F9E11.8"  
/notes="predicted by Genemark.hmm"  
complement(join(<23758..23877,23959..24226,24316..24511,  
24596..24709,24847..25011,25104..25256,25334..25486,  
25730..25897,26036..26185,26555..>26726))  
/gene="F9E11.8"  
complement(join(23758..23877,23959..24226,24316..24511,  
24596..24709,24847..25011,25104..25256,25334..25486,  
25730..25897,26036..26185,26555..26726))  
/gene="F9E11.8"  
/codon\_start=1  
/product="hypothetical protein; 26726-23758"

mRNA  
complement(join(<23758..23877,23959..24226,24316..24511,  
24596..24709,24847..25011,25104..25256,25334..25486,  
25730..25897,26036..26185,26555..26726))  
/gene="F9E11.8"

gene  
complement(join(23758..23877,23959..24226,24316..24511,  
24596..24709,24847..25011,25104..25256,25334..25486,  
25730..25897,26036..26185,26555..26726))  
/gene="F9E11.8"

Query Match 87.0%; Score 17.4; DB 15; Length 39601;  
Best Local Similarity 94.7%; Pred. No. 1.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAACAG 19  
|||||||  
DB 1623 CAAGTAAATGCAGACAG 1605  
|||||||

RESULT 50  
AC020579/c  
LOCUS  
DEFINITION Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence,  
complete sequence.  
ACCESSION AC020579  
VERSION AC020579.5 GI:12324896  
KEYWORDS HTG.  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 50821)  
Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J.,  
Wu, D., Maiti, R., Ronning, C.M., Koo, H., Fujii, C.Y., Uterback, T.R.,  
Barnstead, M.E., Bowman, C.B., White, O., Nieman, W.C. and Fraser, C.M.  
Arabidopsis thaliana chromosome 1 BAC F1017 genomic sequence  
Unpublished  
REFERENCE 2 (bases 1 to 50821)  
Lin, X. and Kaul, S.  
Direct Submission  
Submitted (05-JAN-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA, xlin@igr.org  
REFERENCE 3 (bases 1 to 50821)

## AUTHORS

Town, C.D. and Kaul, S.

## TITLE

Direct Submission

## JOURNAL

Submitted (19-Jan-2001) The Institute for Genomic Research, 9712

## COMMENT

Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org  
 On Jan 19, 2001 this sequence version replaced gi:12280787.  
 Address all correspondence to: at@tigr.org

BAC clone F1017 is from Arabidopsis thaliana chromosome 1

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of several methods: Gene prediction programs including GenScan+ (Chris Burge, <http://CCR-081.mt.edu/GENSCAN.html>), GeneMarkEM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of Glimmer3, see Mihaela Pertea, <http://www.tigr.org/soflab/glimmerm.htm/glimmerm.html>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

## FEATURES

## source

1. .50821  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3702"  
 /chromosome="1"  
 /clone="F1017"  
 /ecotype="Columbia"  
 /product="putative choline kinase; 11757-10052"  
 /protein\_id="AAG52400.1"  
 /db\_xref="GI:12324898"  
 /translation="MGTEKNYENKQYRLPREVKALQAIASEWEDVIDSKALQVPL  
 KGAMTNEVFOIKWPTKEGPKSKLVRIYEGVEIFFDREDIRTFEPMSKHGQPLL  
 LGRFNGRIEELHARTLSACDLRDEISRIATRMKEPHGEMPGAKALLMDRLN  
 WLTA CKRLASPEAKSFLDVMEMINLEKSLFDNDENIGCHNDLQGNIMDEET  
 KAITIDYSCYNPAYDIANHFCMAADYHTETPHIMDYSKIPEVERQRLKTYM  
 SYSDKPSDTMYKKLEDEVEKYLASHLIWLGWGIISEHVHIDFDYMEYARQFEQY  
 WLTKPRLAASEHK"  
 complement (12470. .12501)  
 /rpt\_family="AT rich"  
 complement (13045. .13067)  
 /rpt\_family="(A)n"  
 13093. .16423  
 /genes="F1017.2"  
 /note="identical to heat shock protein 101 GB:AAF26423  
 from [Arabidopsis thaliana]"  
 join (<13093. .14242, 14327. .14508, 14649. .15274, 15359. .15836,  
 15941. .16423)  
 /genes="F1017.2"  
 join (13093. .14242, 14327. .14508, 14649. .15274, 15359. .15836,  
 15941. .16240)  
 /genes="F1017.2"  
 /codon\_start=1

/product="heat shock protein 101; 13093-16240"  
 /protein\_id="AAG52410.1"  
 /db\_xref="GI:12324908"  
 /translation="MNPEKFTHTNETIATAHELVNAGHAQFTPLHLAALISDPTG  
 IFQAISSAGGENAAQAEVINQALKKLPSPDPDIPASSSLIKVIRRAQAQKS  
 RPDTHLAVDOLIMGLLEDQIIRLLNEVGVTARVKSVEKRGKGVESASGDTN  
 FOALKTYGRDIVEOAGKLDVIGRDEIRRVIRLSRRTKNPVLIGEPGVGKTAVVE  
 GLAORIVKGDVPSLTDVRLISLDMGALVAGAKYRGGEFEERLKSVLKEDEAGKVIL  
 FIDIEHLVIGAGKTEGSDMANLFPKMLARGQLACIGATLLEERKVKYKDAAPERF  
 QOVTVAPSPDPTISILRGLKBEYGHGVRIDQRLALINAQLSARYITGRHLPDKAI  
 DLVDEACANVRVQLDSQPEEDINLERKMQLEIELHALEREKDKASKARLITVRKGL  
 LDRKLOPLTMKYRKEKERIDEIRLLKQREELMFSLQAEARRYDLAAADRLGAIQ  
 EVESAIAQLECTSEENMLTENVPHEIAEWSRWTCIPVTRLQONEKERLIGADR  
 LHKRVQONQAVNVAEAILRSRAGLGRPOQPTGSGFLGPTGVGKTEKALAEOLF  
 DDELLVRIDMSEYMEQSHVSKLIGAPPGYGVHEEGQLTEAVRRRKYCVLFEVEK  
 AHVAVNTLLQVLDGRLTDGQRTGRTVFRNSVITMNLGAELLAGLUGTKVTMEVAR  
 DCVNRVRFKHEPPELLNRDIEIVFDFSLHDQLRKVARLQMKDVAVRLAEARGALAV  
 DAALDYTLAESYDPVYGARPIRRMEKKVTVTELSKMVVREEDENSTVYIDAGAGDLV  
 YRVESGGLVDASTGCKSDVLIHANGPKRSDAAQAVKMRTEIEDDDNEEMIED"  
 complement (16481. .17587)  
 /gene="F1017.3"  
 /note="similar to unknown protein GB:AAD21437 from  
 [Arabidopsis thaliana]"  
 complement (join (<16481. .17251, 17318. .>17587))  
 /gene="F1017.3"  
 complement (join (16481. .17251, 17318. .17587))  
 /gene="F1017.3"  
 /codon\_start=1  
 /product="unknown protein; 17587-16481"  
 /protein\_id="AAG52411.1"  
 /db\_xref="GI:12324909"  
 /translation="MAIGPSSVIIIGIILALTYQSLKPPPNLCSGSPGGPITAPR  
 IKRDXRHLAYKEYPREKAKHIVFIHGSDSRHRDAVFATLLSPDLVQRGVYMWV  
 FDPKGYGESDPPITRTPKSLADIEELADSLGSKFVYIGKSMGQAAMGCKLTYPH  
 RLAVGTVLVAPVVYWRNLPLNI STEGNLQQRDQWVAHVAPMLIYWNNTQNVF  
 PGSSVNRDGGVLSPQDKDIIILKLGSSRKPLAEVRQGIHESINRDMIVGFGNWEFD  
 PLELENPLNREGSVHLWQGDDEMLVPVTLQRYTADKLPLWHLHYEVAGCGHFFPLAKG  
 VVDIIVKTLNH"  
 complement (18687. .21119)  
 /gene="F1017.4"  
 /note="similar to unknown protein GB:AAD21437 from  
 [Arabidopsis thaliana]"  
 complement (join (<18687. .19284, 20351. .20547, 20799. .>21119))  
 /gene="F1017.4"  
 complement (join (18687. .19284, 20351. .20547, 20799. .21119))  
 /gene="F1017.4"  
 /codon\_start=1  
 /product="unknown protein; 21119-18687"  
 /protein\_id="AAG52412.1"  
 /db\_xref="GI:12324910"  
 /translation="WASDSRSKSLKPSRSKFLPSPVIVIVIGIIVALTQSKLKPP  
 QPKLGGSSGPPITAPRIKLQDGRYLAYKEHGLPREKANRKIVFIHGSDDCNHDAVFA  
 TLLSPDLVVEELGVYMWSPDRPGYCESDPHSPRTPRSLVSDIEELDQLSLGSKFYVIG  
 KSMGQAAMGCKNLKYI PHRLAGVTLVAPVYVYWRNLPLNVSTEGFNFQQRDQWV  
 RVAVHAPMLIYWNNTQNVFPGSSIANRDSLLSQSDRDIISKRGYTRKPHAEVRQGI  
 HESINRDMIVGFGNWEFDPLDNPFLNNEGFLHWQGDDEMLVPVKLQRYLAHLPLW  
 VHYHSPVRSKGFHFHTKGVVDNIVITLLTTDTDIRQ"  
 complement (21829. .23197)  
 /gene="F1017.5"  
 /note="similar to unknown protein GB:AAD21437 from  
 [Arabidopsis thaliana]"  
 complement (join (<21829. .22626, 22877. .>23197))  
 /gene="F1017.5"  
 complement (join (21829. .22626, 22877. .23197))  
 /gene="F1017.5"  
 /codon\_start=1  
 /product="unknown protein; 23197-21829"  
 /protein\_id="AAG52399.1"  
 /db\_xref="GI:12324897"  
 /translation="WTRDSSRNSSLKMPWRKFLPSPVIVIVIGIIVALTQSKLKPP  
 PPKLGGSSGPPITAPRIKLQDGRYLAYKEHGLPREKANRKIVFIHGSDDCNHDAVFA  
 TLLSPDLVVEELGVYMWSPDRPGYCESDPHSPRTPRSLVSDIEELDQLSLGSKFYVIG  
 YSMGQAAMGCKLYI PHRLAGVTLVAPVYVYWRNLPLNVSTEGFNFQQRDQWV



gene AHYTPWLIYWNWTKWFFGSSIANRHSLLAQPKDIIISKLSSRKPWHAEVROQGIH  
ESINRDMIVGFNGWEPGLDLENPLFNKSGSVHLWQGDMLVPKQLQRYLAHQLPWV  
HYHEVPRSGHFHYHTKGVVDDIVKSLTSDVRSRVCII  
23489..24554  
/gene="F1017.6"

mRNA /note="similar to ribosomal protein L33B GB:NP\_014877 from  
[Saccharomyces cerevisiae]"  
join(<23489..23520,23750..23785,24176..24404,24499..24554)  
/gene="F1017.6"

CDS join(23489..23520,23750..23785,24176..24404,24499..24540)  
/gene="F1017.6"

/codon\_start=1  
/product="putative ribosomal protein; 23489-24540"  
/protein\_id="AAGS2401.1"  
/db\_xref="GI:12324899"  
/translation="MVKGRGRLVYVGTILGVKRSKNOYPTSLVQIRGVNTOE  
ENVYKGRGRLVYVGTILGVKRSKNOYPTSLVQIRGVNTOE  
VRVFMVPSNI  
25509..29950

gene /gene="F1017.7"  
/note="similar to GB:K1AA0361 from [Homo sapiens]"  
join(<25509..25911,25995..26563,26663..29047,  
29144..29950)  
/gene="F1017.7"

CDS join(25509..25911,25995..26563,26663..29047,29144..29950)  
/gene="F1017.7"  
/codon\_start=1  
/product="putative phosphoribosylformylglycinamide  
synthase; 25509-29950"  
/protein\_id="AAGS2403.1"  
/db\_xref="GI:12324901"

Query Match 87.0%; Score 17.4; DB 15; Length 50821;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAG 19

Db 47324 CAAGTAAATGCAGACAG 47306

RESULT 51  
LOCUS AC019512 51405 bp DNA linear HTG 03-JAN-2000  
DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*.  
AC019512  
VERSION AC019512.1 GI:6665385  
KEYWORDS HTG; HTGS PHASE2.  
SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 51405)

REFERENCE Adams M. and Venter, J.C.  
AUTHORS Direct Submission  
TITLE Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
JOURNAL  
COMMENT This sequence was identified as CDM:10210037 by the submitter.  
For more information on this record e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

FEATURES  
source Location/Qualifiers

1..51405  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:722"

ORIGIN

Query Match 87.0%; Score 17.4; DB 14; Length 51405;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAGTAAATGCAGAACAG 20

Db 18370 AGGTAATGCAGAACAG 18388

RESULT 52  
LOCUS AC100240/c

DEFINITION Mus musculus clone RP23-64E10, LOW-PASS SEQUENCE SAMPLING.  
AC100240  
VERSION AC100240.1 GI:17047606  
KEYWORDS HTG; HTGS PHASE0.  
SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 54288)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Mus musculus, clone RP23-64E10  
JOURNAL Unpublished  
REFERENCES 2 (bases 1 to 54288)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepel, Y., Collangelo, M., Collins, S., Collamore, A., Cook, A.,  
Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
Jones, C., Kamat, A., Karatas, A., Kells, C., LaRocque, K.,  
Lamazares, R., Landers, T., Lechoczy, J., Levine, R., Liu, G.,  
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
McCarthy, M., McSwan, P., McKernan, K., McPheeters, R., Meldrin, J.,  
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Olivier, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
TITLE Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L14437  
Center clone name: 64\_E\_10  
-----

\* NOTE: This record contains 68 individual

\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 694: contig of 694 bp in length

\* 695 794: gap of 100 bp  
\* 795 1507: contig of 713 bp in length  
\* 1508 1507: gap of 100 bp  
\* 1608 2279: contig of 672 bp in length  
\* 2280 2379: gap of 100 bp  
\* 3088 3087: contig of 708 bp in length  
\* 3188 3187: gap of 100 bp  
\* 3898 3897: contig of 710 bp in length  
\* 3998 4705: contig of 708 bp in length  
\* 4706 4805: gap of 100 bp  
\* 4806 5490: contig of 685 bp in length  
\* 5491 5590: gap of 100 bp  
\* 5591 6282: contig of 692 bp in length  
\* 6283 6382: gap of 100 bp  
\* 6383 7088: contig of 706 bp in length  
\* 7089 7188: gap of 100 bp  
\* 7189 7871: contig of 683 bp in length  
\* 7872 7971: gap of 100 bp  
\* 7972 8651: contig of 680 bp in length  
\* 8652 8751: gap of 100 bp  
\* 8752 9460: contig of 709 bp in length  
\* 9461 9560: gap of 100 bp  
\* 9561 10265: contig of 705 bp in length  
\* 10266 11085: gap of 100 bp  
\* 11086 11887: contig of 720 bp in length  
\* 11887 11887: contig of 702 bp in length  
\* 11888 11987: gap of 100 bp  
\* 11988 12688: contig of 701 bp in length  
\* 12689 12788: gap of 100 bp  
\* 12789 13485: contig of 697 bp in length  
\* 13486 13585: gap of 100 bp  
\* 13586 14268: contig of 683 bp in length  
\* 14269 14368: gap of 100 bp  
\* 14369 15078: contig of 710 bp in length  
\* 15079 15178: gap of 100 bp  
\* 15179 15890: contig of 712 bp in length  
\* 15891 15990: gap of 100 bp  
\* 15991 16692: contig of 702 bp in length  
\* 16693 16792: gap of 100 bp  
\* 16793 17465: contig of 673 bp in length  
\* 17466 17565: gap of 100 bp  
\* 17566 18237: contig of 672 bp in length  
\* 18238 18337: gap of 100 bp  
\* 18338 19029: contig of 692 bp in length  
\* 19030 19129: gap of 100 bp  
\* 19130 19828: contig of 699 bp in length  
\* 19829 19928: gap of 100 bp  
\* 19929 20610: contig of 682 bp in length  
\* 20611 20710: gap of 100 bp  
\* 20711 21415: contig of 705 bp in length  
\* 21416 21515: gap of 100 bp  
\* 21516 22224: contig of 709 bp in length  
\* 22225 22324: gap of 100 bp  
\* 22325 23024: contig of 700 bp in length  
\* 23025 23124: gap of 100 bp  
\* 23125 23825: contig of 701 bp in length  
\* 23826 23925: gap of 100 bp  
\* 23926 24628: contig of 703 bp in length  
\* 24629 24728: gap of 100 bp  
\* 24729 25444: contig of 716 bp in length  
\* 25445 25445: gap of 100 bp  
\* 25446 26274: contig of 730 bp in length  
\* 26275 26374: gap of 100 bp  
\* 26375 27066: contig of 692 bp in length  
\* 27067 27166: gap of 100 bp  
\* 27167 27852: contig of 686 bp in length  
\* 27853 27952: gap of 100 bp  
\* 27953 28662: contig of 710 bp in length  
\* 28663 28762: gap of 100 bp  
\* 28763 29463: contig of 701 bp in length  
\* 29464 29563: gap of 100 bp

\* 29564 30266: contig of 703 bp in length  
\* 30267 30366: gap of 100 bp  
\* 30367 31070: contig of 704 bp in length  
\* 31071 31170: gap of 100 bp  
\* 31171 31885: contig of 715 bp in length  
\* 31886 31985: gap of 100 bp  
\* 31986 32669: contig of 684 bp in length  
\* 32670 32769: gap of 100 bp  
\* 32770 33451: contig of 682 bp in length  
\* 33452 33551: gap of 100 bp  
\* 33552 34243: contig of 692 bp in length  
\* 34244 34343: gap of 100 bp  
\* 34344 35047: contig of 704 bp in length  
\* 35048 35147: gap of 100 bp  
\* 35148 35858: contig of 711 bp in length  
\* 35859 35958: gap of 100 bp  
\* 35959 36685: contig of 727 bp in length  
\* 36686 37495: contig of 710 bp in length  
\* 37496 37595: gap of 100 bp  
\* 37596 38294: contig of 699 bp in length  
\* 38295 38394: gap of 100 bp  
\* 38395 39096: contig of 702 bp in length  
\* 39097 39196: gap of 100 bp  
\* 39197 39882: contig of 686 bp in length  
\* 39883 39982: gap of 100 bp  
\* 39983 40665: contig of 683 bp in length  
\* 40666 40765: gap of 100 bp  
\* 40766 41474: contig of 709 bp in length  
\* 41475 41574: gap of 100 bp  
\* 41575 42253: contig of 679 bp in length  
\* 42254 42353: gap of 100 bp  
\* 42354 43048: contig of 695 bp in length  
\* 43049 43148: gap of 100 bp  
\* 43149 43862: contig of 714 bp in length  
\* 43863 43962: gap of 100 bp  
\* 43963 44643: contig of 681 bp in length  
\* 44644 44743: gap of 100 bp  
\* 44744 45461: contig of 718 bp in length  
\* 45462 45561: gap of 100 bp  
\* 45562 46252: contig of 691 bp in length  
\* 46253 46353: gap of 100 bp  
\* 46353 47059: contig of 707 bp in length  
\* 47060 47159: gap of 100 bp  
\* 47160 47864: contig of 705 bp in length  
\* 47865 47964: gap of 100 bp  
\* 47965 48676: contig of 712 bp in length  
\* 48677 48776: gap of 100 bp  
\* 48777 49479: contig of 703 bp in length  
\* 49480 49579: gap of 100 bp  
\* 49580 50279: contig of 700 bp in length  
\* 50280 50379: gap of 100 bp  
\* 50380 51068: contig of 689 bp in length  
\* 51069 51168: gap of 100 bp  
\* 51169 51879: contig of 711 bp in length  
\* 51880 51979: gap of 100 bp  
\* 51980 52689: contig of 710 bp in length  
\* 52690 52789: gap of 100 bp  
\* 52790 53480: contig of 691 bp in length  
\* 53481 53580: gap of 100 bp  
\* 53581 54288: contig of 708 bp in length.

FEATURES source  
Location/Qualifiers 1. 54288

Query Match 87.0%; Score 17.4; DB 14; Length 54288;  
Best Local Similarity 94.7%; Pred. NO. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAACAGG 20  
DB 21776 AAGTAAATGCAGAACAGG 21758



[illegible]

```
* 32610 32709: gap of 100 bp
* 32710 33419: contig of 710 bp in length
* 33420 33519: gap of 100 bp
* 33520 34208: contig of 689 bp in length
* 34209 34308: gap of 100 bp
* 34309 34954: contig of 646 bp in length
* 34955 35054: gap of 100 bp
* 35055 35734: contig of 680 bp in length
* 35735 35834: gap of 100 bp
* 35835 36507: contig of 673 bp in length
* 36508 36607: gap of 100 bp
* 36608 37273: contig of 666 bp in length
* 37274 37373: gap of 100 bp
* 37374 38052: contig of 679 bp in length
* 38053 38152: gap of 100 bp
* 38153 38859: contig of 707 bp in length
* 38860 38959: gap of 100 bp
* 38960 39664: contig of 705 bp in length
* 39665 39764: gap of 100 bp
* 39765 40448: contig of 684 bp in length
* 40449 40548: gap of 100 bp
* 40549 41272: contig of 724 bp in length
* 41273 41373: gap of 100 bp
* 41373 42086: contig of 714 bp in length
* 42087 42186: gap of 100 bp
* 42187 42866: contig of 680 bp in length
* 42867 42966: gap of 100 bp
* 42967 43660: contig of 694 bp in length
* 43661 43760: gap of 100 bp
* 43761 44449: contig of 689 bp in length
* 44450 44549: gap of 100 bp
* 44550 45222: contig of 673 bp in length
* 45223 45322: gap of 100 bp
* 45323 46002: contig of 680 bp in length
* 46003 46102: gap of 100 bp
* 46103 46788: contig of 686 bp in length
* 46789 46888: gap of 100 bp
* 46889 47606: contig of 718 bp in length
* 47607 47706: gap of 100 bp
* 47707 48430: contig of 724 bp in length
* 48431 48530: gap of 100 bp
* 48531 49241: contig of 711 bp in length
* 49242 49341: gap of 100 bp
* 49342 50060: contig of 719 bp in length
* 50061 50160: gap of 100 bp
* 50161 50868: contig of 708 bp in length
* 50869 50968: gap of 100 bp
* 50969 51649: contig of 681 bp in length
* 51650 51749: gap of 100 bp
* 51750 52427: contig of 678 bp in length
* 52428 52527: gap of 100 bp
* 52528 53202: contig of 675 bp in length
* 53203 53302: gap of 100 bp
* 53303 53957: contig of 655 bp in length
* 53958 54057: gap of 100 bp
* 54058 54747: contig of 690 bp in length

Query Match      87.0%; Score 17.4; DB 14; Length 63540;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
    |||||
Db 9262 CAAGTAAATTCAGAAACAG 9244

RESULT 54
AL590286/c
LOCUS
DEFINITION
  Human DNA sequence from clone RP11-157L10 on chromosome 6 Contains
  a ribosomal protein L34 (RPL34) pseudogene, complete sequence.
ACCESSION
  AL590286
VERSION
  AL590286.12 GI:17384111
```

HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Homnidae; Homo.  
1 (bases 1 to 65756)  
Tracey, A.  
Direct Submission  
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On Dec 5, 2001 this sequence version replaced gi:16944115.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em.; EMBL; Sw.; SWISSPROT; Tr.; TREMBL; Wp.; WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr6  
RP11-157L10 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VSCITOR: pBACE3.6  
----- Genome Center  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: vegas@sanger.ac.uk  
-----  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one subclone; and the assembly was confirmed by restriction digest,  
except on the rare occasion of the clone being a YAC.  
Location/Qualifiers  
1. .65756  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="RP11-157L10"  
/clone\_lib="RPCI-11.1"  
2000  
/note="Clone right end: RP11-57O22"  
complement(6159..6449)  
/locus\_tag="RP11-157L10.1-001"  
/pseudogene  
complement(6159..6449)  
/locus\_tag="RP11-157L10.1-001"  
/note="match: proteins: AAL87001 P49207 Q29223 Q9BUZ2  
Q9D1R9"  
/pseudogene  
complement(6159..6449)  
/locus\_tag="RP11-157L10.1-001"  
/note="Clone left end: RP11-62E18"  
63757  
/note="Clone left end: RP1-257A15"

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
misc\_feature  
gene  
CDS  
misc\_feature  
misc\_feature  
ORIGIN  
Query Match 87.0%; Score 17.4; DB 8; Length 65756;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 AAGTAAATGCAGAAACAG 20

Db	537		537	AAGTAAATACAGAACAGG	519	
RESULT 55	AC003096/c					
LOCUS	AC003096					
DEFINITION	Arabidopsis thaliana chromosome 2 clone T29F13 map ve016, complete sequence.					
ACCESSION	AC003096					
VERSION	AC003096.3					
KEYWORDS	HTG.					
SOURCE	Arabidopsis thaliana (thale cress)					
ORGANISM	Arabidopsis thaliana					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.					
REFERENCE	1 (bases 1 to 69817)					
AUTHORS	Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C., Sykes,S.M., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D., Somerville,C.R. and Venter,J.C.					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 69817)					
AUTHORS	Lin,X.					
TITLE	Direct Submission					
JOURNAL	Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA					
REFERENCE	3 (bases 1 to 69817)					
AUTHORS	Town,C.D. and Kaul,S.					
TITLE	Direct Submission					
JOURNAL	Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org					
COMMENT	On Apr 18, 2002 this sequence version replaced gi:6598389.					
FEATURES	Location/Qualifiers					
source	1. .69817					
	/organism="Arabidopsis thaliana"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:3702"					
	/chromosome="2"					
	/map="ve016"					
	/clone="T29F13"					
	/ecotype="Columbia"					
misc_feature	complement(1..1207)					
gene	/notes="overlap with BAC clone F1913 (AC004238:1..1207)"					
	complement(4397..6665)					
	/gene="At2g34770"					
	/notes="synonym: T29F13.2; identical to GB:AF021804; supported by cDNA: gi_15215595_gb_AY050326.1"					
mRNA	complement(join(<4397..4589,4692..4838,4916..5021,5155..5351,5460..5552,5705..5770,6641..6665))					
	/gene="At2g34770"					
CDS	complement(join(4470..4589,4692..4838,4916..5021,5155..5351,5460..5552,5705..5755))					
	/gene="At2g34770"					
	/codon_start=1					
	/product="fatty acid hydroxylase (FAH1)"					
	/protein_id="AAC16270.1"					
	/db_xref="GI:3132481"					
	/translation="MWAQGFVDLKKPLVQVGHGEDVEVWHQFIATKEGRRFFQS DFWEFLTVMAVPVIMLVWVVCISRSVSMGCSLPEIPIVMGIFITWFFEYVLH RFVEFHITKSYNGNTAHYLIHGCHKHPMDHLVFPPTAILCPFWNIAKAISTP STAPALEGGMLGVMDVTHYLLHQAQPTPTVTKNKKYHLNHHFRIQDKGFGITSS LWDVFGTLPTITKAPKEQ"					
	7682..7862					
gene	/note="synonym: T29F13.3; pseudogene, tubulin beta chain"					
	/pseudo					
	8739..12082					
gene	/gene="At2g34750"					
	/note="synonym: T29F13.4"					
mRNA	join(<8739..8867,8989..9078,9171..9242,9347..9418,9505..9741,9833..9979,10031..10135,10253..10378,10487..10555,10667..10788,11071..11158,11340..11453,11547..11731,11830..>12082)					
	/gene="At2g34750"					
	/codon_start=1					
	/product="unknown protein"					
	/protein_id="AAC16259.1"					
	/db_xref="GI:3132470"					
	/translation="MGAVELMSDPSSLCTVENYVDNVLSDTQLVQTVTKALTSVKTG DSDLYSENVGMWARDIKFEPKQPDVVAQLETLVKALSGAVACIDVLHKKLSALPMK LWDHSPVDALNVLNVLATVTSGLKLSCLNMLVSNFVPPVQVNNVSHSLANKKI DVLSPVHAALLKISILVPLTSPRLVPLFQQPMKHKHDSIIVIVYESSLKLENSIG QVGGSMIIQWYMERLDRDVRQNNLIQLRSLIEDDDIPQDSSRGKGMFDMLEDAE GTMDGDLPGVPLKQDTSIGSVKSLDKLMLVAVFHEHLESQNGRLQDVFEESLRLI FENFLNTYKSKFTQFLIFYACSLDPENGVKFAKLVBIPLSSNKHVATROASLRLI DEVCYCTCDDTRPEAHQIFPFGCOALMVYLCFMRSLIDVPRFSOLTPLESILM HKLNPLMVCLPSVVAEFURQAKGEGFIVSDSFIQDLSLSRAFGFPERLDTFFP FDPCLLSKNSGIFSPNFIYGNWKATYDEDDDDNDAEVIUNGDESDSDEDEADLDLAL NKMSITPKHSFKNMERDLRLMPSRIIRPSTSPESL"					
	complement(11877..11962)					
repeat_region	/rpt_family="CAT'n"					
gene	complement(12226..13210)					
	/gene="At2g34740"					
mRNA	/note="synonym: T29F13.5"					
	complement(join(<12226..12375,12458..12612,12713..12989,13073..13210))					
	/gene="At2g34740"					
CDS	/codon_start=1					
	/product="putative protein phosphatase 2C"					
	/protein_id="AAC16260.1"					
	/db_xref="GI:3132471"					
	/translation="MEDFIVADTKTVKGNHGLYAIIPDGSSGSDVADYLVNHLFDNIL SQDFWRNPKKAIKRAYKSTDDYILQNVVPRGGSTAVTAIVDGKKIIVANVGDGRA ILCRSDVVQITVDHEPKRDLVSKGKGFVSKPGNVPVVDGLAMTRAFDGGGLK EHSVTPNIEIAEIHDDTKFLILASDGLWKVMSNDEVDQIKKRGNAEAAKMLIDKA LARGSKDDISCVVVSFLQWID"					
repeat_region	13276..13359					
repeat_region	/rpt_family="CAT'n"					
gene	complement(14252..14307)					
	/rpt_family="AT rich"					
	complement(15165..17920)					
	/gene="At2g34730"					
mRNA	/note="synonym: T29F13.6"					
	complement(join(<15165..15344,15503..15629,15738..>17920))					
	/gene="At2g34730"					
CDS	complement(join(15165..15344,15503..15629,15738..17920))					
	/gene="At2g34730"					
	/codon_start=1					
	/product="putative myosin heavy chain"					
	/protein_id="AAC16261.1"					
	/db_xref="GI:3132472"					
	/translation="MEEVYKNGSLEFHDRTLSSSLQVNGVLKENENPDVDFLEDL DSYWEDINRLTISRVVSDIIRGMVTAIESDAAEKIAQKDLSEKIRETLLLYHVGS ENESESHIDDELTDQSSSLKKARQQLMLVLELTNLREVIHNGSGATVDOSL GDSSPHETRSKTVDLSDLSILKLVLRKNDMLPSWQEHDFQKEIESAVTS VLRLKDEYEDLLDQKAEFGGNSLIIGNIKETIGLROELAEIRKSLFDHENGDEAG EVGDKRVRQLHRKMSGLNSVSSWNGKHESSTGLIPEHNETHRMSPDENIHF KIEMNKMKDHDYKIOELTEQCFPKRYKLYNLTERGFSFGVKDKELGALKKIPFI SKLDLKMEDEKPFVSEKNDAGLRKOLDLLENROLKDSLSAAEKMSLSQAEADH QELIRKLTDEDSRNEASIEDVYCGFVFTEFVCGQKCTQETDLEHSLMREAYELL EDLARKEARKSKEDFEDSCVKSVMMECCSVIYKEAHEAKKIIVLNHVTKEGTL RSEMVDKERLKEIHRGLVKENLVQTAENNLATERKKIIVSQIINDLSQVER QETEIQDKTEALSVWSARELVKVPYETKISSIRELELAKSKEMKQKQSKSLSL SETKASETKLKKQLSLDVLVPPQLIKGFDILEGLIAEKTOKTNSRLKNWQSQSDLS HQINEVKGKASTYKQLEKCCDLKKAEBVDLGDVEVTLDDLLEKIYIALDHYSPI LKHYPGVSTISVYTFIFISILRLIMRPY"					
	19787..19824					
repeat_region						

```
repeat_region
gene
  /rpt_family="AT rich"
  complement(20248..20271)
  /rpt_family="(A)n"
  21380..22513
  /gene="At2g34720"
  /note="synonym: T29F13.7"
  join(<21380..21511,21666..21761,21897..21980,22069..22233,
  22310..>22513)
  /gene="At2g34720"
  join(21380..21511,21666..21761,21897..21980,22069..22233,
  22310..22513)
  /gene="At2g34720"
  /codon_start=1
  /product="putative CCAAT-binding transcription factor
  subunit"
  /protein_id="AAC16262.1"
  /db_xref="GI:3132473"
  /translation="MPCWSDNNESHAKRPDSOTRPOVPSGRSSSEIDTWSVSEPM
  PCFDAYNAKTLWFLIIOAKSRDRLSAVLDPAGHLYPPDPYRSVFAQAQYLPHPY
  PVQIQMGMOQPGPLOCDAVEEPVFNAKQYHGLRLRQSRKLEARNRAIKAKP
  YMHESRHLHATRRPGCGRFLNAKKENGDKHKEBEATSDENTSEASSLRSEKLA
  TSGPNGRS"
  2437..24462
  /rpt_family="AT rich"
  complement(26798..26850)
  /rpt_family="AT rich"
  complement(27849..27880)
  /rpt_family="(A)n"
  27932..33247
  /note="molecular marker Athb_14"
  28199..33208
  /gene="At2g34710"
  /note="synonym: T29F13.8; identical to GB:Y11122"
  join(28199..28770,29333..29492,29584..29699,29785..29875,
  29992..30087,30169..30366,30441..30596,30681..30755,
  30844..30950,31036..31225,31317..31505,31597..31677,
  31760..31900,31992..32305,32404..32518,32604..32753,
  32837..33208)
  /gene="At2g34710"
  complement(28213..28274)
  /rpt_family="(GGAG)n"
  join(28538..28770,29333..29492,29584..29699,29785..29875,
  29992..30087,30169..30366,30441..30596,30681..30755,
  30844..30950,31036..31225,31317..31505,31597..31677,
  31760..31900,31992..32305,32404..32518,32604..32753,
  32837..32983)
  /gene="At2g34710"
  /codon_start=1
  /product="homeodomain transcription factor (ATHB-14)"
  /db_xref="GI:3132474"
  /translation="MMVHMSRDMNRRSPDKGLDSGKYRVYRTPQEVALERVYTEC
  PKPSLRRLQRLRCPILSNIPKQVFNRRCKRORKEARLQTVNRKLNAMNK
  LLMEENDRLQKOVSNLYENGHKQLHTASCTTTDNCSESVVSGQQHQPKNPQH
  QORDANNPAGLSIAEAEAEFLSKATGATDWNQVGNKPGPDSIGVIAIRNCSGI
  AARACGLVLEPMKVABILDKRPSRLDCRSVDTLSVPAGNGTIELITYTOMYAPT
  LAARDFTWLRYSCTLEDGSYVVCPSRLTSATGPTGPPSNFVRAEMKPSGLIRPC
  DGGSTLHIVDVLDAVSPVPLVYESSKILAKQMTVAALRHVROIQAETSQVQ
  YGGGQPAVLRTFSORLCRPFNDVNGFVDDGWSPMGSDGAEVTVMLNLSPGKFGS
  QVGNFSLPSFGSGVLCAKASMLLQNVPPAVLVRFLRHRSEWADVGVDAYAAASLRAS
```

```
DEFINITION
  Human DNA sequence from clone RP11-309C8 on chromosome X, complete
  sequence.
ACCESSION
  AL662897
VERSION
  AL662897.6
GI:18307360
KEYWORDS
  HTG.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
  Hominidae; Homo.
REFERENCE
  1. (bases 1 to 84987)
  Howden, P.
  Direct Submission
  Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
  Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
  Clone requests: clonerequest@sanger.ac.uk
  On Jan 23, 2002 this sequence version replaced gi:18250572.
  The following abbreviations are used to associate primary accession
  numbers given in the feature table with their source databases:
  Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
  on the WORMPEP database can be found at
  http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
  was generated from part of bacterial clone contigs of human
  chromosome X, constructed by the Sanger Centre Chromosome X Mapping
  Group. Further information can be found at
  http://www.sanger.ac.uk/HGP/ChrX
  RP11-309C8 is from the library RPCI-11.2 constructed by the group
  of Pieter de Jong. For further details see
  http://www.chori.org/bacpac/home.htm
  VECTOR: pBACe3.6
  ----- Genome Center
  Center: Wellcome Trust Sanger Institute
  Center code: SC
  Web site: http://www.sanger.ac.uk
  Contact: vegas@sanger.ac.uk
  -----
  This sequence was finished as follows unless otherwise noted: all
  regions were either double-stranded or sequenced with an alternate
  chemistry or covered by high quality data (i.e., phred quality >=
  30); an attempt was made to resolve all sequencing problems, such
  as compressions and repeats; all regions were covered by at least
  one subclone; and the assembly was confirmed by restriction digest,
  except on the rare occasion of the clone being a YAC.
  location/Qualifiers
  1..84987
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /clone="RP11-309C8"
  /clone_lib="RPCI-11.2"
  -----
  QUERY MATCH
  Query Match 87.0%; Score 17.4; DB 8; Length 84987;
  Best Local Similarity 94.7%; Pred. No. 1.2e+03;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
  QY 1 CAAGTAAATGCAGAAACAG 19
  ||||| ||||| ||||| |||||
  DB 64774 CAAGTTAATGCAGAAACAG 64756
  -----
  RESULT 57
  AC106751
  LOCUS
  DEFINITION
  Homo sapiens chromosome 5 clone RP11-164E14, WORKING DRAFT
  HTG 12-JAN-2002
  AC106751
  SEQUENCE, 3 unrounded pieces.
  AC106751
  AC106751.1 GI:18139301
  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
  SOURCE
  Homo sapiens (human)
  ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
```

```

Hominidae; Homo.
REFERENCE
1 (bases 1 to 101526)
AUTHORS
DOE Joint Genome Institute.
TITLE
Sequencing of Human Chromosome 5
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 101526)
AUTHORS
DOE Joint Genome Institute.
TITLE
Direct Submission
JOURNAL
Submitted (12-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 469451
Center clone name: RPCI-11_164E14
-----
Summary Statistics
Consensus quality: 100219 bases at least Q40
Consensus quality: 101000 bases at least Q30
Consensus quality: 101186 bases at least Q20
Estimated insert size: 90000; agarose-fp estimation
Estimated insert size: 101326; sum-of-contigs estimation
Quality coverage: 13.95 in Q20 bases; agarose-fp estimation
Quality coverage: 12.39 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1345: contig of 1345 bp in length
* 1346 1445: gap of unknown length
* 1446 45004: contig of 43559 bp in length
* 45005 45104: gap of unknown length
* 45105 101526: contig of 56422 bp in length.
FEATURES
source
1..101526
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-164E14"
/clone_lib="RPCI human BAC library 11"
1346..1445
/estimated_length=unknown
45005..45104
/estimated_length=unknown
ORIGIN
Query Match 87.0%; Score 17.4; DB 14; Length 101526;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAAGTAAATGCAGAAACAG 19
|||||
Db 12535 CAAGTAAAGTCAGAAACAG 12553
RESULT 58
AL136124 104632 bp DNA linear PRI 18-MAY-2005
LOCUS
Human DNA sequence from clone RP4-792D7 on chromosome 1q42.2-43
DEFINITION
Contains the 5' end of the TARBP1 gene for TAR (HIV) RNA binding
protein 1 and a CpG island, complete sequence.
ACCESSION
AL136124
VERSION
AL136124.10 GI:9662905
KEYWORDS
HTG; TARBP1.
SOURCE
Homo sapiens (human)

```

---

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 (bases 1 to 104632)
AUTHORS
Martin, S.
TITLE
Direct Submission
JOURNAL
Submitted (13-MAY-2005) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries: vegas@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Aug 2, 2000 this sequence version replaced gi:9211777.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep/
This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: vegas@sanger.ac.uk
-----
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats, all regions were covered by at least
one subclone; and the assembly was confirmed by restriction digest,
except on the rare occasion of the clone being a YAC.
RP4-792D7 is from the library RPCI-4 constructed by the group of
Piet de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.
Location/Qualifiers
1..104632
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704D07792"
/db_xref="taxon:9606"
/chromosome="1"
/map="q42.2-43"
/clone="RP4-792D7"
/clone_lib="RPCI-4"
misc_feature
1
/note="Clone left end: RP4-792D7"
gene
Join(complement(59831..60761),complement(54372..54469),
complement(52846..52915),complement(49160..49308),
complement(47314..47366),complement(45495..45592),
complement(41919..42054),complement(40888..40984),
complement(39325..39414),complement(32086..32224),
complement(30138..30237),complement(28461..28633),
complement(18933..19030),complement(15068..15229),
complement(11745..11959),complement(11074..11352),
complement(10794..10982),complement(9234..9424),
complement(8975..9046),complement(7331..7451),
complement(2360..2470),
complement(AL355472.19:90370..90519),
complement(AL355472.19:82731..82817),
complement(AL355472.19:78189..78385),
complement(AL355472.19:73467..73548),
complement(AL355472.19:70668..70839),
complement(AL355472.19:65932..66123),
complement(AL355472.19:65648..65772),
complement(AL355472.19:64702..64838),
complement(AL355472.19:63599..64031))
/locus_tag="RP5-827C21.5-001"
Join(complement(59831..60761),complement(54372..54469),
complement(52846..52915),complement(49160..49308),
complement(47314..47366),complement(45495..45592),

```

```
complement (41919. .42054), complement (40888. .40984),
complement (39325. .39414), complement (32086. .32224),
complement (30138. .30237), complement (28461. .28633),
complement (18933. .19030), complement (15068. .15229),
complement (11745. .11959), complement (11074. .11335),
complement (10794. .10982), complement (9234. .9424),
complement (8975. .9046), complement (7331. .7451),
complement (2360. .2470),
complement (AL355472.19:90370. .90519),
complement (AL355472.19:82731. .82817),
complement (AL355472.19:78189. .78385),
complement (AL355472.19:73467. .73548),
complement (AL355472.19:70668. .70839),
complement (AL355472.19:65932. .66123),
complement (AL355472.19:65648. .65772),
complement (AL355472.19:64702. .64838),
complement (AL355472.19:63599. .64031),
/locus_tag="RP5-827C21.5-001",
/product="TAR (HIV) RNA binding protein 1"
/note="match: ESTs: Em:AA714641.1 Em:AI079700.1
Em:BM989347.1
match: cDNAs: Em:U38847.1"
join(complement (59831. .60761), complement (54372. .54469),
complement (52846. .52915), complement (49160. .49308),
complement (47314. .47366), complement (45495. .45592),
complement (41919. .42054), complement (40888. .40984),
complement (39325. .39414), complement (32086. .32224),
complement (30138. .30237), complement (28461. .28633),
complement (18933. .19030), complement (15068. .15229),
complement (11745. .11959), complement (11074. .11335),
complement (10794. .10982), complement (9234. .9424),
complement (8975. .9046), complement (7331. .7451),
complement (2360. .2470),
complement (AL355472.19:90370. .90519),
complement (AL355472.19:82731. .82817),
complement (AL355472.19:78189. .78385),
complement (AL355472.19:73467. .73548),
complement (AL355472.19:70668. .70839),
complement (AL355472.19:65932. .66123),
complement (AL355472.19:65648. .65772),
complement (AL355472.19:64702. .64838),
complement (AL355472.19:63863. .64031),
/locus_tag="RP5-827C21.5-001",
/standard_name="OTHUMP00000036048"
/note="match: proteins: Tr:Q13395"
/codon_start=1
/product="TAR (HIV) RNA binding protein 1"
/protein_id="CAI22828.1"
/db_xref="GI:56204383"
/db_xref="Genew:11568"
/db_xref="GOA:Q13395"
/db_xref="InterPro:IPR001537"
/db_xref="UniProt/TrEMBL:Q13395"
translation="NEWVLAELLSQSDPRALLGALCGEASAEVETLFLQLRLE
DEAGSGGAGALPAANEVAGYVPLRLSRGPAGDPDSLPQRRHRLVRAAGA
ALRSCVRLAGVLPQALAAEALRDLLAGWAPGAEEVLAAGPCLRPREDGFLI
ERVATVAGLALGGGDDGDEAGPAEDAAALVAGRLPVLCQCGAALRVAGGLAAG
ASLGSRYVEKLLVLSALAEKLLPEPGDRAGAEAGPDRRCWRFWRTQAGLQGA
DALTKRYALQRAVEVSAELGADCTGPGNGPSLFPWSEKDKLLKFWENYIL
TNETLEGQIHWIKVPLKLNLFYAVSENGCWLPHPSWHMCYKRFSENKILS
KGVYTHFLYETKILPSPSEFIIGPLMDALSESLYSRSPQPTGSCSPGLGLK
KQGLVYLSLPEETKSFLLKFIRKTSRHWCAVPFLFLSKALANVRHKAIGDLG
LARDVHICTMTHTQLLRGAQCYLLQTMALLNDEKVSLSVDSFTLMSRQESLG
RGTSLWELCDMLRNESYFKPSPTCSSTGLKHTSLNAYKVSIVQEVYKSAWETGEN
CMPWDFEAKVLSMLVLAADVDEGMKTOYSGKQRTENVLRIPLDLLVMKFSINAY
MPLLTDLRCLQLLLKLLNLCRLKGSSAQDDDEVSTVLQNFPMSTTESIQHILRLTMN
ELNSVDDLRKHLNMLTELNLHNLKWKGNFPIWRVLSLLKNASISOHLQMSGQ
EPTQSYQORVVSMAALAVCEAIQDKPELQDLSHAGLESLFLSLQNLQTLQKPHA
EFTQSGVPLECSLESSSQGKIVQYIHQWVLCFLFKKXHTLIPITGSEI
LEPFLVQMPNRTLQSALEALTVLSDOVLPVFLKLVLPKILTSSELICIESFDM
ANKWITSSNTOLIFWANKAFVQVFNKVLITIAAKIKGOAYFKIKIMYKIMBSA
IKTGVPNTLISYCCQSWTVSASNVQSGLSSAKNYSLEILACIFGTGFRDORLVQD
VOTFIENGHDCANIVMENTKREDHYRICAVKFLCLLDGNSMHSKLFIEDLAIKLL
```

CDS

```
DKDLVSKKRYVYVNSLOHRVKNRVQWTLVLPRLDQNFNGIIDIIRFOAGFTNNO
ASIKYFIEMIIILILHKFQPLKFWDFSVGBENLATSICTFLAVLSHLDIITQINIP
EKKILQALIVLQWCFNFRFVRLYALVALKKLWTVCKVLSVEEDALTPVLESSL
HOVSEMHGAGNKKNNWQRIQEHFFKATFHPKDKCYLETIFILPRLSGLIEDWITID
KFRFTDVPPLAAGFQWYLSQTSKSLKPGDMSQDQDIGNLVNADNQAEWTDVQKKIIP
WNSRVSLDLLELLFQDRAARLKGKISRLIVVASLIDKPTNLGGLCRTEVFGASVLVV
GSLQICSDKQFHLVSVAEQWLPVAVKPPQLIDYLOQKKTEGYTIIGVEQTAKSLDL
TQCFPEKSLLLGNREGIPANLIQQLDVCVEIPQGGIIRSLNVHVSGLLIWEYTR
QQLUSHGDTKP"
join(complement (2360. .2472),
complement (AL355472.19:90370. .90519),
complement (AL355472.19:82690. .82817))
/locus_tag="RP5-827C21.5-004"
join(complement (2360. .2472),
complement (AL355472.19:90370. .90519),
complement (AL355472.19:82690. .82817))
/locus_tag="RP5-827C21.5-004"
/product="TAR (HIV) RNA binding protein 1"
/note="match: ESTs: Em:AA836178.1"
misc_feature
104632
/note="Clone_right_end: RP4-792D7"
ORIGIN
Query Match 87.0%; Score 17.4; DB 8; Length 104632;
Best Local Similarity 94.7%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 AAGTAAATGCAGAAACAGG 20
|||||
Db 22543 AAGTAAATGCAGACAGG 22561
|||||
RESULT 59
CR543861_32/c
WPCOMMENT
Sequence split into 36 fragments LOCUS CR543861 Accession CR543861
Fragment Name Begin End
CR543861_00 1 110000
CR543861_01 100001 210000
CR543861_02 200001 310000
CR543861_03 300001 410000
CR543861_04 400001 510000
CR543861_05 500001 610000
CR543861_06 600001 710000
CR543861_07 700001 810000
CR543861_08 800001 910000
CR543861_09 900001 1010000
CR543861_10 1000001 1110000
CR543861_11 1100001 1210000
CR543861_12 1200001 1310000
CR543861_13 1300001 1410000
CR543861_14 1400001 1510000
CR543861_15 1500001 1610000
CR543861_16 1600001 1710000
CR543861_17 1700001 1810000
CR543861_18 1800001 1910000
CR543861_19 1900001 2010000
CR543861_20 2000001 2110000
CR543861_21 2100001 2210000
CR543861_22 2200001 2310000
CR543861_23 2300001 2410000
CR543861_24 2400001 2510000
CR543861_25 2500001 2610000
CR543861_26 2600001 2710000
CR543861_27 2700001 2810000
CR543861_28 2800001 2910000
CR543861_29 2900001 3010000
CR543861_30 3000001 3110000
CR543861_31 3100001 3210000
CR543861_32 3200001 3310000
CR543861_33 3300001 3410000
CR543861_34 3400001 3510000
CR543861_35 3500001 3598621
```

Continuation (33 of 36) of CR543861 from base 3200001 (CR543861 Acinetobacter sp. ADP1 d

Query Match 87.0%; Score 17.4; DB 1; Length 110000;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
Db 38781 CAAGTAAATGCAGAAACTG 38763

RESULT 60  
CR848719 2  
WPCOMMENT

Sequence split into 4 fragments LOCUS CR848719 Accession CR848719  
Fragment Name Begin End  
CR848719\_0 1 110000  
CR848719\_1 100001 210000  
CR848719\_2 200001 310000  
CR848719\_3 300001 353035

Continuation (3 of 4) of CR848719 from base 200001 (CR848719 Danio rerio clone DKEY-269D

Query Match 87.0%; Score 17.4; DB 14; Length 110000;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
|||||  
Db 18237 AAGAAATGCAGAAACAGG 18255

Search completed: January 27, 2006, 23:14:09  
Job time : 793.944 secs

CR543861

***This Page Blank (uspto)***



Result No.	Score	Query Match	Length	DB	ID	Description
1	23	100.0	23	14	ADZ75861	Adz75861 Group B s
2	23	100.0	1731	6	ABN68276	Abn68276 Streptoco
3	23	100.0	1734	13	ADV85417	Adv85417 Streptoco
4	23	100.0	29072	13	ADV87723	Adv87723 Streptoco
5	23	100.0	29072	13	ADV78976	Adv78976 Streptoco
6	23	100.0	110000	6	ABN71527_07	Continuation (8 of
7	23	100.0	110000	13	ADH81201_08	Continuation (9 of
8	19.8	86.1	1378	10	ADD68798	Add68798 Streptoco
9	19.8	86.1	1378	12	ADJ62002	Adj62002 Group A s
10	19.8	86.1	1379	10	ADD68800	Add68800 Streptoco
11	19.8	86.1	1379	10	ADD68797	Add68797 Streptoco
12	19.8	86.1	1379	10	ADD68791	Add68791 Streptoco
13	19.8	86.1	1379	12	ADJ62001	Adj62001 Group A s
14	19.8	86.1	1379	12	ADJ62004	Adj62004 Group A s
15	19.8	86.1	1379	12	ADJ61995	Adj61995 Group A s
16	19.8	86.1	1384	10	ADD68792	Add68792 Streptoco
17	19.8	86.1	1384	10	ADD68794	Add68794 Streptoco
18	19.8	86.1	1384	12	ADJ61998	Adj61998 Group A s
19	19.8	86.1	1384	12	ADJ61996	Adj61996 Group A s

PA (REIS/) REISER A.  
 XX  
 PI Uhl JR, Cockerill FR, Aichinger C, Reiser A;  
 XX WPI; 2005-371550/38.  
 XX  
 DR  
 PT  
 PT  
 PT  
 XX  
 XX  
 PS Claim 2; SEQ ID NO 1; 13pp; English.  
 XX  
 XX The invention relates to a real-time PCR-based method of detecting the  
 CC presence or absence of group B streptococcus (GBS) bacterial pathogens in  
 CC a biological sample from an individual. The method comprises  
 CC amplification of a conserved region of the phosphotransferase (pts) gene  
 CC of GBS (especially using PCR primers AD275861-AD275862), detection of the  
 CC amplification product with a pair of fluorescently labeled  
 CC phosphotransferase probes (especially AD275863-AD275864), and detection  
 CC of the presence or absence of fluorescence resonance energy transfer  
 CC (FRET), where presence of FRET is indicative of the presence of GBS in  
 CC the sample. GBS infection is a leading cause of neonatal morbidity and  
 CC mortality, with infection occurring during childbirth. Currently, it is  
 CC recommended that women are screened for GBS during week 35-37 of  
 CC gestation by a culture-based method which may take up to 72 hours for a  
 CC result. However, many women first present at healthcare facilities at the  
 CC time of labor, and in addition, GBS infection can be transient, so that a  
 CC woman free of GBS at the time of screening may not be free of GBS when  
 CC she is due to give birth. The method of the invention provides a real-  
 CC time assay for the detection of group B streptococcus in a sample, is  
 CC more sensitive and specific than prior art non-culture based methods, and  
 CC can thus be implemented for routine diagnosis of the presence of group B  
 CC streptococcus. Sequences AD275861-AD275862 represent GBS  
 CC phosphotransferase PCR primers specifically claimed for use in the method  
 CC of the invention.  
 XX  
 SQ Sequence 23 BP; 9 A; 2 C; 8 G; 4 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 23; DB 14; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 DB 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 RESULT 2  
 ID ABN68276  
 XX ABN68276 standard; DNA; 1731 BP.  
 XX  
 AC ABN68276;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polynucleotide SEQ ID NO 4465.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX  
 PF 29-OCT-2001; 2001WO-GB004789.  
 XX  
 XX 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX

PA (CHIR-) CHIRON SPA.  
 XX (GENO-) INST GENOMIC RES.  
 PI Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;  
 PI Tetelin H;  
 XX  
 DR WPI; 2002-352536/38.  
 DR P-P8DB; ABP27645.  
 XX  
 PT New Streptococcus protein for the treatment or prevention of infection or  
 PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 XX  
 XX Claim 7; Page 3607; 4525pp; English.  
 XX  
 CC The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX  
 SQ Sequence 1731 BP; 541 A; 323 C; 373 G; 494 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 23; DB 6; Length 1731;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 |||||  
 DB 177 TGAGAAGGCAGTAGAAGCTTAG 199  
 |||||  
 RESULT 3  
 ID ADV85417  
 XX ADV85417 standard; DNA; 1734 BP.  
 XX  
 AC ADV85417;  
 XX  
 DT 24-FEB-2005 (first entry)  
 XX  
 DE Streptococcus agalactiae DNA sequence, SEQ ID 6558.  
 XX  
 KW Antibacterial; vaccine; bacterial infection; ds.  
 XX  
 OS Streptococcus agalactiae.  
 XX  
 PN WO200292818-A2.  
 XX  
 PD 21-NOV-2002.  
 XX  
 PF 26-APR-2002; 2002WO-IB003059.  
 XX  
 PR 26-APR-2001; 2001FR-00005642.  
 XX  
 PA (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.  
 DR

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
PT and identification of therapeutic targets.  
XX  
XX Claim 4; SEQ ID NO 6558; 439pp; French.  
XX  
XX The present invention relates to novel Streptococcus agalactiae  
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transporters, cell membrane proteins and  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial S. agalactiae infection. The complete  
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.  
XX  
XX Sequence 1734 BP; 540 A; 323 C; 376 G; 495 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 23; DB 13; Length 1734;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 177 TGAGAAGGCAGTAGAAGCTTAG 199  
RESULT 4  
ADV87723  
ID ADV87723 standard; DNA; 29072 BP.  
XX  
AC ADV87723;  
XX  
XX 24-FEB-2005 (first entry)  
XX  
XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
XX  
XX Antibacterial; Vaccine; bacterial infection; ds.  
XX  
XX Streptococcus agalactiae.  
XX  
XX FR2824074-A1.  
XX  
XX 31-OCT-2002.  
XX  
XX 26-APR-2001; 2001FR-00005642.  
XX  
XX 26-APR-2001; 2001FR-00005642.  
XX  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
XX WPI; 2004-101891/11.  
XX  
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX PT and identification of therapeutic targets.  
XX  
XX Claim 1; SEQ ID NO 117; 2687pp; French.  
XX  
XX The present invention relates to novel Streptococcus agalactiae

CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
CC agalactiae involved in the synthesis of amino acids, cell membranes,  
CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
CC phospholipid metabolism, nucleotide metabolism including purines,  
CC pyrimidines and/or nucleosides, regulatory functions, replication,  
CC transcription, translation, protein transport, adaptation to atypical  
CC conditions, sensitivity to medicines and/or analogues, functions related  
CC to transporters, cell membrane proteins and cellular machinery. (I) are  
CC useful for the detection and/or amplification of nucleic acids.  
CC Pharmaceutical composition comprising (I) or (II) are useful for  
CC treatment of a bacterial S. agalactiae infection. Note: WO200222818A2 is  
CC equivalent for the present basic patent FR2824074A1. WO200222818A2  
CC contains 6617 sequence whereas the present patent only contains 2344  
CC sequences.  
XX  
XX Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 23; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 12072 TGAGAAGGCAGTAGAAGCTTAG 12094  
RESULT 5  
ADV78976  
ID ADV78976 standard; DNA; 29072 BP.  
XX  
AC ADV78976;  
XX  
XX 24-FEB-2005 (first entry)  
XX  
XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
XX  
XX Antibacterial; vaccine; bacterial infection; ds.  
XX  
XX Streptococcus agalactiae.  
XX  
XX WO200222818-A2.  
XX  
XX 21-NOV-2002.  
XX  
XX 26-APR-2002; 2002WO-IB003059.  
XX  
XX 26-APR-2001; 2001FR-00005642.  
XX  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
XX WPI; 2004-101891/11.  
XX  
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX PT and identification of therapeutic targets.  
XX  
XX Claim 1; SEQ ID NO 117; 439pp; French.  
XX  
XX The present invention relates to novel Streptococcus agalactiae  
CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transporters, cell membrane proteins and  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial S. agalactiae infection. The complete  
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.  
XX  
XX Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 23; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 2.9;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 12072 TGAGAAGGCAGTAGAAGCTTAG 12094  
RESULT 5  
ADV78976  
ID ADV78976 standard; DNA; 29072 BP.  
XX  
AC ADV78976;  
XX  
XX 24-FEB-2005 (first entry)  
XX  
XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
XX  
XX Antibacterial; vaccine; bacterial infection; ds.  
XX  
XX Streptococcus agalactiae.  
XX  
XX WO200222818-A2.  
XX  
XX 21-NOV-2002.  
XX  
XX 26-APR-2002; 2002WO-IB003059.  
XX  
XX 26-APR-2001; 2001FR-00005642.  
XX  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
XX WPI; 2004-101891/11.  
XX  
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX PT and identification of therapeutic targets.  
XX  
XX Claim 1; SEQ ID NO 117; 439pp; French.  
XX  
XX The present invention relates to novel Streptococcus agalactiae

CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial S. agalactiae infection. The complete  
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.

XX  
SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

Query Match 100.0%; Score 23; DB 13; Length 29072;

Best Local Similarity 100.0%; Pred. No. 2.9; Mismatches 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23

Db 12072 TGAGAAGGCAGTAGAAGCTTAG 12094

#### RESULT 6

ABN71527\_07  
Continuation (8 of 22) of ABN71527 from base 700001 (Streptococcus polynucleotide SEQ ID  
WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527

Fragment Name	Begin	End
WP ABN71527_00	1	110000
WP ABN71527_01	100001	210000
WP ABN71527_02	200001	310000
WP ABN71527_03	300001	410000
WP ABN71527_04	400001	510000
WP ABN71527_05	500001	610000
WP ABN71527_06	600001	710000
WP ABN71527_07	700001	810000
WP ABN71527_08	800001	910000
WP ABN71527_09	900001	1010000
WP ABN71527_10	1000001	1110000
WP ABN71527_11	1100001	1210000
WP ABN71527_12	1200001	1310000
WP ABN71527_13	1300001	1410000
WP ABN71527_14	1400001	1510000
WP ABN71527_15	1500001	1610000
WP ABN71527_16	1600001	1710000
WP ABN71527_17	1700001	1810000
WP ABN71527_18	1800001	1910000
WP ABN71527_19	1900001	2010000
WP ABN71527_20	2000001	2110000
WP ABN71527_21	2100001	2155561

Query Match 100.0%; Score 23; DB 6; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23

Db 90942 TGAGAAGGCAGTAGAAGCTTAG 90964

#### RESULT 7

ADV81204\_08  
Continuation (9 of 23) of ADV81204 from base 800001 (Streptococcus agalactiae complete g  
WP Sequence split into 23 fragments LOCUS ADV81204 Accession ADV81204

Fragment Name	Begin	End
WP ADV81204_00	1	110000
WP ADV81204_01	100001	210000
WP ADV81204_02	200001	310000
WP ADV81204_03	300001	410000
WP ADV81204_04	400001	510000
WP ADV81204_05	500001	610000
WP ADV81204_06	600001	710000
WP ADV81204_07	700001	810000
WP ADV81204_08	800001	910000
WP ADV81204_09	900001	1010000
WP ADV81204_10	1000001	1110000

WP	ADV81204_11	1100001	1210000
WP	ADV81204_12	1200001	1310000
WP	ADV81204_13	1300001	1410000
WP	ADV81204_14	1400001	1510000
WP	ADV81204_15	1500001	1610000
WP	ADV81204_16	1600001	1710000
WP	ADV81204_17	1700001	1810000
WP	ADV81204_18	1800001	1910000
WP	ADV81204_19	1900001	2010000
WP	ADV81204_20	2000001	2110000
WP	ADV81204_21	2100001	2210000
WP	ADV81204_22	2200001	2217924

Query Match 100.0%; Score 23; DB 13; Length 110000;

Best Local Similarity 100.0%; Pred. No. 3.3; Mismatches 0; Indels 0; Gaps 0;  
Matches 23; Conservative 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23

Db 71277 TGAGAAGGCAGTAGAAGCTTAG 71299

#### RESULT 8

ADD68798  
ID ADD68798 standard; DNA; 1378 BP.

XX ADD68798;

AC

XX 15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' ptaI DNA - SEQ ID 14.

XX beta-haemolytic Group A Streptococcus; GAS;

XX fluorescence resonance energy transfer; FRET; ptaI; ds.

XX Streptococcus sp. 'group A'.

XX US6593093-B1.

XX 15-JUL-2003.

XX 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2003-828265/77.

PT Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.

XX Example 2; SEQ ID NO 14; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridising step comprising contacting the  
CC amplified product with fluorescent-labelled probes and finally, detecting  
CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 14)  
CC of the invention.

XX Sequence 1378 BP; 414 A; 289 C; 308 G; 367 T; 0 U; 0 Other;

```

DE Streptococcus sp. 'group A' pteI DNA - SEQ ID 16.
XX
XX beta-haemolytic Group A Streptococcus; GAS;
KW fluorescence resonance energy transfer; FRET; pteI; ds.
XX
XX Streptococcus sp. 'group A'.
XX
XX US6593093-B1.
XX
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
XX Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
XX Example 2; SEQ ID NO 16; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridising step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' pteI DNA (SEQ ID 16)
XX of the invention.
XX
XX SQ Sequence 1379 BP; 420 A; 291 C; 307 G; 361 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1379;
XX Best Local Similarity 91.3%; Pred. No. 58;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0
XX
XX QY 1 TGAGAGGCAGTAGAAGCTTAG 23
XX |||||
XX DB 10 TGAATGCAGTAGAAGCTTAG 32
XX |||||
XX
XX RESULT 11
XX ADD68797
XX ID ADD68797 standard; DNA; 1379 BP.
XX
XX AC ADD68797;
XX
XX DT 15-JAN-2004 (first entry)
XX
XX DE Streptococcus sp. 'group A' pteI DNA - SEQ ID 13.
XX
XX KW beta-haemolytic Group A Streptococcus; GAS;
KW fluorescence resonance energy transfer; FRET; pteI; ds.
XX
XX OS Streptococcus sp. 'group A'.
XX
XX US6593093-B1.
XX
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX PF 20-FEB-2002; 2002US-00081923.
XX
XX
XX
XX

```

```
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
XX Example 2; SEQ ID NO 13; 35pp; English.
XX
CC The invention relates to a novel method for detecting the presence or
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
CC isolated from an individual by performing more than one cycling step. The
CC method involves an amplifying step comprising contacting the sample with
CC GAS primers, a subsequent hybridising step comprising contacting the
CC amplified product with fluorescent-labelled probes and finally, detecting
CC fluorescence resonance energy transfer (FRET) where the presence or
CC absence of FRET indicates the presence or absence of GAS in the sample.
CC The method of the invention may be useful for detecting the presence or
CC absence of GAS in a biological sample from an individual. The current
CC sequence is that of the Streptococcus sp. 'group A' ptsl DNA (SEQ ID 13)
CC of the invention.
XX
SQ Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;

Query Match      86.1%; Score 19.8; DB 10; Length 1379;
Best Local Similarity 91.3%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTACAGAAAGCTTAG 23
Db 9 TGAAAATGCAGTACAGAAAGCTTAG 31

RESULT 12
ADD68791
ID ADD68791 standard; DNA; 1379 BP.
XX
AC ADD68791;
XX
XX 15-JAN-2004 (first entry)
XX
DE Streptococcus sp. 'group A' ptsl DNA - SEQ ID 7.
XX
KW beta-haemolytic Group A Streptococcus; GAS;
KW fluorescence resonance energy transfer; FRET; ptsl; ds.
XX
OS Streptococcus sp. 'group A'.
XX
PN US6593093-B1.
XX
PD 15-JUL-2003.
XX
PF 20-FEB-2002; 2002US-00081923.
XX
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
```

```
PS Example 2; SEQ ID NO 7; 35pp; English.
XX
CC The invention relates to a novel method for detecting the presence or
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
CC isolated from an individual by performing more than one cycling step. The
CC method involves an amplifying step comprising contacting the sample with
CC GAS primers, a subsequent hybridising step comprising contacting the
CC amplified product with fluorescent-labelled probes and finally, detecting
CC fluorescence resonance energy transfer (FRET) where the presence or
CC absence of FRET indicates the presence or absence of GAS in the sample.
CC The method of the invention may be useful for detecting the presence or
CC absence of GAS in a biological sample from an individual. The current
CC sequence is that of the Streptococcus sp. 'group A' ptsl DNA (SEQ ID 7)
CC of the invention.
XX
SQ Sequence 1379 BP; 415 A; 289 C; 309 G; 366 T; 0 U; 0 Other;

Query Match      86.1%; Score 19.8; DB 10; Length 1379;
Best Local Similarity 91.3%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTACAGAAAGCTTAG 23
Db 9 TGAAAATGCAGTACAGAAAGCTTAG 31

RESULT 13
ADD62001
ID ADJ62001 standard; DNA; 1379 BP.
XX
AC ADJ62001;
XX
XX 06-MAY-2004 (first entry)
XX
DE Group A streptococcus ptsl gene #8.
XX
KW Group A Streptococcus; GAS; gene; ds.
XX
OS Streptococcus sp.
XX
PN US2004014118-A1.
XX
PD 22-JAN-2004.
XX
PF 19-JUN-2003; 2003US-00465205.
XX
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
DR WPI; 2004-224229/21.
XX
PT Detecting Group A Streptococcus in biological sample, involves contacting
PT sample with pair of ptsl primers and probes and detecting fluorescence
PT resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 13; 36pp; English.
XX
CC The invention relates to a method of detecting the presence or absence of
CC Group A Streptococcus (GAS) in biological sample from individual, by
CC contacting sample with pair of ptsl primers and probes to produce ptsl
CC amplification product, where first, second probes are labelled with donor
CC and acceptor fluorescent moiety, and detecting presence or absence of
CC fluorescence resonance energy transfer, which indicates the presence or
CC absence of GAS in biological sample. The method is useful for detecting
CC the presence or absence of Group A Streptococcus in a biological sample
CC such as throat swabs, tissues and bodily fluids from an individual. The
CC present sequence represents a group A streptococcus ptsl gene.
XX
XX Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;
```

```
Query Match      86.1%; Score 19.8; DB 12; Length 1379;
Best Local Similarity 91.3%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGGAGGCGAGTGAAGGCTTAG 23
DB 9 TGAATAATGCAGTAGAAGGCTTAG 31

RESULT 14
ADJ62004
ID ADJ62004 standard; DNA; 1379 BP.
XX
XX
AC ADJ62004;
XX
XX 06-MAY-2004 (first entry)
XX
XX Group A streptococcus PtsI gene #11.
XX
XX Group A Streptococcus; GAS; gene; ds.
XX
XX Streptococcus sp.
XX
XX US2004014118-A1.
XX
XX 22-JAN-2004.
XX
XX 19-JUN-2003; 2003US-00465205.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptsI primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 16; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptsI primers and probes to produce ptsI
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptsI gene.
XX
XX Sequence 1379 BP; 420 A; 291 C; 307 G; 361 T; 0 U; 0 Other;

Query Match      86.1%; Score 19.8; DB 12; Length 1379;
Best Local Similarity 91.3%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGGAGGCGAGTGAAGGCTTAG 23
DB 10 TGAATAATGCAGTAGAAGGCTTAG 32

RESULT 15
ADJ61995
ID ADJ61995 standard; DNA; 1379 BP.
XX
XX
AC ADJ61995;
XX
XX 06-MAY-2004 (first entry)
XX
XX
```

```
DE Group A streptococcus PtsI gene #2.
XX
XX Group A Streptococcus; GAS; gene; ds.
XX
XX Streptococcus sp.
XX
XX US2004014118-A1.
XX
XX 22-JAN-2004.
XX
XX 19-JUN-2003; 2003US-00465205.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptsI primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 7; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptsI primers and probes to produce ptsI
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptsI gene.
XX
XX Sequence 1379 BP; 415 A; 289 C; 309 G; 366 T; 0 U; 0 Other;

Query Match      86.1%; Score 19.8; DB 12; Length 1379;
Best Local Similarity 91.3%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGGAGGCGAGTGAAGGCTTAG 23
DB 9 TGAATAATGCAGTAGAAGGCTTAG 31

RESULT 16
ADD68792
ID ADD68792 standard; DNA; 1384 BP.
XX
XX ADD68792;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptsI DNA - SEQ ID 8.
XX
XX beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptsI; ds.
XX
XX Streptococcus sp. 'group A'.
XX
XX US6593093-B1.
XX
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
```

```

PI Uhl JR, Cockerill FR;
XX WPI; 2003-828265/77.
XX
PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
XX Example 2; SEQ ID NO 8; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridizing step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptsl DNA (SEQ ID 8)
XX
XX SQ Sequence 1384 BP; 421 A; 291 C; 306 G; 366 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1384;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
DB 10 TGAATAATGCAGTAGAAGCTTAG 32
XX
XX RESULT 17
XX ADD68794
XX ID ADD68794 standard; DNA; 1384 BP.
XX AC
XX AC ADD68794;
XX
XX DT 15-JAN-2004 (first entry)
XX
XX DE Streptococcus sp. 'group A' ptsl DNA - SEQ ID 10.
XX
XX KW beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptsl; da.
XX
XX OS Streptococcus sp. 'group A'.
XX
XX PN US6593093-B1.
XX
XX PD 15-JUL-2003.
XX
XX PF 20-FEB-2002; 2002US-00081923.
XX
XX PR 20-FEB-2002; 2002US-00081923.
XX
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX PI Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
XX Detecting presence/absence of group A Streptococcus in sample by
XX performing cycling step comprising amplifying, hybridizing DNA from
XX sample by fluorescent probe, detecting for fluorescent resonance energy
XX transfer.
XX
XX Example 2; SEQ ID NO 10; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridizing step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptsl DNA (SEQ ID 8)
XX
XX SQ Sequence 1384 BP; 416 A; 292 C; 308 G; 368 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1384;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
DB 9 TGAATAATGCAGTAGAAGCTTAG 31
XX
XX RESULT 18
XX ADJ61998
XX ID ADJ61998 standard; DNA; 1384 BP.
XX
XX AC ADJ61998;
XX
XX DT 06-MAY-2004 (first entry)
XX
XX DE Group A streptococcus ptsl gene #5.
XX
XX KW Group A Streptococcus; GAS; gene; da.
XX
XX OS Streptococcus sp.
XX
XX PN US2004014118-A1.
XX
XX PD 22-JAN-2004.
XX
XX PF 19-JUN-2003; 2003US-00465205.
XX
XX PR 20-FEB-2002; 2002US-00081923.
XX
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX PI Uhl JR, Cockerill FR;
XX
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptsl primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 10; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptsl primers and probes to produce ptsl
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptsl gene.
XX
XX SQ Sequence 1384 BP; 416 A; 292 C; 308 G; 368 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 12; Length 1384;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX

```



QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 DB 9 TGAATAATGCAGTAGAAGCTTAG 31

## RESULT 19

ADJ61996  
 ID ADJ61996 standard; DNA; 1384 BP.  
 AC ADJ61996;  
 DT 06-MAY-2004 (first entry)  
 DE Group A streptococcus PtsI gene #3.  
 KW Group A Streptococcus; GAS; gene; ds.  
 OS Streptococcus sp.  
 PN US2004014118-A1.  
 XX 22-JAN-2004.

PF 19-JUN-2003; 2003US-00465205.  
 PR 20-FEB-2002; 2002US-00081923.  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 PI Uhl JR, Cockerill FR;  
 XX WPI; 2004-224229/21.

PT Detecting presence/absence of Group A Streptococcus in biological sample, involves contacting sample with pair of ptsI primers and probes and detecting fluorescence resonance energy transfer indicating Group A Streptococcus in sample.

PS Disclosure; SEQ ID NO 8; 36pp; English.

XX The invention relates to a method of detecting the presence or absence of Group A Streptococcus (GAS) in biological sample from individual, by contacting sample with pair of ptsI primers and probes to produce ptsI amplification product, where first, second probes are labelled with donor and acceptor fluorescent moiety, and detecting presence or absence of fluorescence resonance energy transfer, which indicates the presence or absence of GAS in biological sample. The method is useful for detecting the presence or absence of Group A Streptococcus in a biological sample such as throat swabs, tissues and bodily fluids from an individual. The present sequence represents a group A streptococcus ptsI gene.

XX Sequence 1384 BP; 421 A; 291 C; 306 G; 366 T; 0 U; 0 Other;

Query Match 86.1%; Score 19.8; DB 12; Length 1384;  
 Best Local Similarity 91.3%; Pred. No. 59;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 DB 10 TGAATAATGCAGTAGAAGCTTAG 32

## RESULT 20

ADJ68793  
 ID ADJ68793 standard; DNA; 1385 BP.

AC ADJ68793;

DT 15-JAN-2004 (first entry)

DE Streptococcus sp. 'group A' ptsI DNA - SEQ ID 9.

KW beta-haemolytic Group A Streptococcus; GAS;  
 KW fluorescence resonance energy transfer; FRET; ptsI; ds.

XX Streptococcus sp. 'group A'.  
 OS US6593093-B1.  
 PN 15-JUL-2003.  
 XX 20-FEB-2002; 2002US-00081923.  
 XX 20-FEB-2002; 2002US-00081923.  
 PR (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX Uhl JR, Cockerill FR;  
 PI WPI; 2003-828265/77.  
 DR Detecting presence/absence of group A Streptococcus in sample by performing cycling step comprising amplifying, hybridizing DNA from sample by fluorescent probe, detecting for fluorescent resonance energy transfer.

XX Example 2; SEQ ID NO 9; 35pp; English.

XX The invention relates to a novel method for detecting the presence or absence of beta-haemolytic Group A Streptococcus (GAS) in a sample isolated from an individual by performing more than one cycling step. The method involves an amplifying step comprising contacting the sample with GAS primers, a subsequent hybridising step comprising contacting the amplified product with fluorescent-labelled probes and finally, detecting fluorescence resonance energy transfer (FRET) where the presence or absence of FRET indicates the presence or absence of GAS in the sample. The method of the invention may be useful for detecting the presence or absence of GAS in a biological sample from an individual. The current sequence is that of the Streptococcus sp. 'group A' ptsI DNA (SEQ ID 9) of the invention.

XX Sequence 1385 BP; 416 A; 293 C; 309 G; 367 T; 0 U; 0 Other;

Query Match 86.1%; Score 19.8; DB 10; Length 1385;  
 Best Local Similarity 91.3%; Pred. No. 59;  
 Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
 DB 10 TGAATAATGCAGTAGAAGCTTAG 32

## RESULT 21

ADJ61997  
 ID ADJ61997 standard; DNA; 1385 BP.

AC ADJ61997;

DT 06-MAY-2004 (first entry)

DE Group A streptococcus PtsI gene #4.

XX Group A Streptococcus; GAS; gene; ds.

OS Streptococcus sp.

PN US2004014118-A1.

PD 22-JAN-2004.

XX 19-JUN-2003; 2003US-00465205.

PR 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2004-224229/21.  
XX Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptsi primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX Disclosure; SEQ ID NO 9; 36pp; English.  
XX The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptsi primers and probes to produce ptsi  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of  
CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a group A streptococcus ptsi gene.  
XX  
SQ Sequence 1385 BP; 416 A; 293 C; 309 G; 367 T; 0 U; 0 Other;  
Query Match 86.1%; Score 19.8; DB 12; Length 1385;  
Best Local Similarity 91.3%; Pred. No. 59;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TGAGAAGCGAGTGAAGGCTTAG 23  
Db 10 TGAATAATGCAGTAGAAGGCTTAG 32  
RESULT 22  
ADD68795  
ID ADD68795 standard; DNA; 1390 BP.  
XX AC  
XX ADD68795;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Streptococcus sp. 'group A' ptsi DNA - SEQ ID 11.  
XX  
KW beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptsi; ds.  
XX  
OS Streptococcus sp. 'group A'.  
XX  
XX US6593093-B1.  
XX  
XX 15-JUL-2003.  
XX  
XX 20-FEB-2002; 2002US-00081923.  
XX  
XX 20-FEB-2002; 2002US-00081923.  
XX  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
XX Uhl JR, Cockerill PR;  
XX  
XX WPI; 2003-828265/77.  
XX  
XX Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.  
XX  
XX Example 2; SEQ ID NO 11; 35pp; English.  
XX  
XX The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridizing step comprising contacting the  
CC amplified product with fluorescent-labelled probes and finally, detecting

CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the Streptococcus sp. 'group A' ptsi DNA (SEQ ID 11)  
CC of the invention.  
XX  
SQ Sequence 1390 BP; 419 A; 293 C; 310 G; 368 T; 0 U; 0 Other;  
Query Match 86.1%; Score 19.8; DB 10; Length 1390;  
Best Local Similarity 91.3%; Pred. No. 59;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TGAGAAGCGAGTGAAGGCTTAG 23  
Db 24 TGAATAATGCAGTAGAAGGCTTAG 46  
RESULT 23  
ADJ61999  
ID ADJ61999 standard; DNA; 1390 BP.  
XX AC  
XX ADJ61999;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
XX Group A streptococcus Ptsi gene #6.  
DE  
XX Group A Streptococcus; GAS; gene; ds.  
KW  
XX Streptococcus sp.  
OS  
XX US2004014118-A1.  
XX  
XX 22-JAN-2004.  
XX  
XX 19-JUN-2003; 2003US-00465205.  
XX  
XX 20-FEB-2002; 2002US-00081923.  
XX  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
XX Uhl JR, Cockerill PR;  
XX  
XX WPI; 2004-224229/21.  
XX  
XX Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptsi primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX  
XX Disclosure; SEQ ID NO 11; 36pp; English.  
XX  
XX The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptsi primers and probes to produce ptsi  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of  
CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a group A streptococcus ptsi gene.  
XX  
SQ Sequence 1390 BP; 419 A; 293 C; 310 G; 368 T; 0 U; 0 Other;  
Query Match 86.1%; Score 19.8; DB 12; Length 1390;  
Best Local Similarity 91.3%; Pred. No. 59;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TGAGAAGCGAGTGAAGGCTTAG 23  
Db 24 TGAATAATGCAGTAGAAGGCTTAG 46

```

RESULT 24
ADD68799
ID ADD68799 standard; DNA; 1393 BP.
XX
XX AC ADD68799;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptai DNA - SEQ ID 15.
XX
XX beta-haemolytic Group A Streptococcus; GAS;
KW fluorescence resonance energy transfer; FRET; ptai; ds.
XX
XX Streptococcus sp. 'group A'.
OS
XX US6593093-B1.
PN
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
PF
XX
XX 20-FEB-2002; 2002US-00081923.
PR
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
PA
XX Uhl JR, Cockerill FR;
PI
XX WPI; 2003-828265/77.
XX
XX Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
XX Example 2; SEQ ID NO 15; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridizing step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 15)
XX of the invention.
XX
XX SQ Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 10; Length 1393;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

The invention relates to a novel method for detecting the presence or
absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
isolated from an individual by performing more than one cycling step. The
method involves an amplifying step comprising contacting the sample with
GAS primers, a subsequent hybridizing step comprising contacting the
amplified product with fluorescent-labelled probes and finally, detecting
fluorescence resonance energy transfer (FRET) where the presence or
absence of FRET indicates the presence or absence of GAS in the sample.
The method of the invention may be useful for detecting the presence or
absence of GAS in a biological sample from an individual. The current
sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 15)
of the invention.

Query Match 86.1%; Score 19.8; DB 10; Length 1393;
Best Local Similarity 91.3%; Pred. No. 59;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 TGAGAGGCGCAGTAGAAGCTTAG 23
10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 25
ADJ62003
ID ADJ62003 standard; DNA; 1393 BP.
XX
XX AC ADJ62003;
XX
XX 06-MAY-2004 (first entry)
XX
XX Group A streptococcus PtaI gene #10.
DE
XX Group A Streptococcus; GAS; gene; ds.
XX
XX

```

```

OS Streptococcus sp.
XX
XX US2004014118-A1.
XX
XX 22-JAN-2004.
XX
XX 19-JUN-2003; 2003US-00465205.
XX
XX 20-FEB-2002; 2002US-00081923.
PR
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
PA
XX Uhl JR, Cockerill FR;
PI
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
PT sample with pair of ptai primers and probes and detecting fluorescence
PT resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 15; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptai primers and probes to produce ptai
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptai gene.
XX
XX SQ Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;
XX
XX Query Match 86.1%; Score 19.8; DB 12; Length 1393;
XX Best Local Similarity 91.3%; Pred. No. 59;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 TGAGAGGCGCAGTAGAAGCTTAG 23
10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 26
ADD68801
ID ADD68801 standard; DNA; 1447 BP.
XX
XX AC ADD68801;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptai DNA - SEQ ID 17.
DE
XX beta-haemolytic Group A Streptococcus; GAS;
KW fluorescence resonance energy transfer; FRET; ptai; ds.
XX
XX Streptococcus sp. 'group A'.
OS
XX US6593093-B1.
PN
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
PF
XX
XX 20-FEB-2002; 2002US-00081923.
PR
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
PA
XX Uhl JR, Cockerill FR;
PI
XX WPI; 2003-828265/77.
XX
XX

```

PT Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.

XX  
XX Example 2; SEQ ID NO 17; 35pp; English.

PS  
XX The invention relates to a novel method for detecting the presence or  
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
XX isolated from an individual by performing more than one cycling step. The  
XX method involves an amplifying step comprising contacting the sample with  
XX GAS primers, a subsequent hybridizing step comprising contacting the  
XX amplified product with fluorescent-labelled probes and finally, detecting  
XX fluorescence resonance energy transfer (FRET) where the presence or  
XX absence of FRET indicates the presence or absence of GAS in the sample.  
XX The method of the invention may be useful for detecting the presence or  
XX absence of GAS in a biological sample from an individual. The current  
XX sequence is that of the Streptococcus sp. 'group A' p<sub>tsi</sub> DNA (SEQ ID 17)  
XX of the invention.

XX  
XX Sequence 1447 BP; 443 A; 304 C; 318 G; 382 T; 0 U; 0 Other;

SEQ Query Match 86.1%; Score 19.8; DB 10; Length 1447;  
Best Local Similarity 91.3%; Pred. No. 59;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||  
DB 10 TGAATAATGCAGTAGAAGCTTAG 32  
|||||

RESULT 27  
ADJ62005  
ID ADJ62005 standard; DNA; 1447 BP.

XX  
XX AC ADJ62005;

XX  
XX DT 06-MAY-2004 (first entry)

XX  
XX DE Group A streptococcus PtsI gene #12.

XX  
XX KW Group A Streptococcus; GAS; gene; ds.

XX  
XX OS Streptococcus sp.

XX  
XX PN US2004014118-A1.

XX  
XX PD 22-JAN-2004.

XX  
XX PF 19-JUN-2003; 2003US-00465205.

XX  
XX PR 20-FEB-2002; 2002US-00081923.

XX  
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX  
XX PI Uhl JR, Cockerill FR;

XX  
XX DR WPI; 2004-224229/21.

XX  
XX PT Detecting Group A Streptococcus in biological sample, involves contacting  
XX sample with pair of p<sub>tsi</sub> primers and probes and detecting fluorescence  
XX resonance energy transfer indicating Group A Streptococcus in sample.

XX  
XX PS Disclosure; SEQ ID NO 17; 36pp; English.

XX  
XX The invention relates to a method of detecting the presence or absence of  
XX Group A Streptococcus (GAS) in biological sample from individual, by  
XX contacting sample with pair of p<sub>tsi</sub> primers and probes to produce p<sub>tsi</sub>  
XX amplification product, where first, second probes are labelled with donor  
XX and acceptor fluorescent moiety, and detecting presence or absence of  
XX fluorescence resonance energy transfer, which indicates the presence or  
XX absence of GAS in biological sample. The method is useful for detecting  
XX the presence or absence of Group A Streptococcus in a biological sample

CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a group A streptococcus p<sub>tsi</sub> gene.

XX  
XX Sequence 1447 BP; 443 A; 304 C; 318 G; 382 T; 0 U; 0 Other;

SEQ Query Match 86.1%; Score 19.8; DB 12; Length 1447;  
Best Local Similarity 91.3%; Pred. No. 59;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||  
DB 10 TGAATAATGCAGTAGAAGCTTAG 32  
|||||

RESULT 28  
ABN68277  
ID ABN68277 standard; DNA; 1731 BP.

XX  
XX AC ABN68277;

XX  
XX DT 01-JUL-2002 (first entry)

XX  
XX DE Streptococcus polynucleotide SEQ ID NO 4467.

XX  
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.

XX  
XX OS Streptococcus pyogenes.

XX  
XX PN WO200234771-A2.

XX  
XX PD 02-MAY-2002.

XX  
XX PF 29-OCT-2001; 2001WO-GB004789.

XX  
XX PR 27-OCT-2000; 2000GB-00026333.

XX  
XX PR 24-NOV-2000; 2000GB-00028727.

XX  
XX PR 07-MAR-2001; 2001GB-00005640.

XX  
XX PA (CHIR-) CHIRON SPA.

XX  
XX PI Telford J, Massignani V, Margarit Y RosI, Grandi G, Fraser C;  
XX Tettelin H;

XX  
XX DR WPI; 2002-352536/38.

XX  
XX DR P-PSDB; ABP27646.

XX  
XX PT New Streptococcus protein for the treatment or prevention of infection or  
XX disease caused by Streptococcus bacteria, such as meningitis, and for  
XX detecting a compound that binds to the protein.

XX  
XX PS Claim 7; Page 3608; 4525pp; English.

XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX (Streptococcus pyogenes), comprising one of 5483 sequences (s1), given in  
XX the specification. The proteins have antibacterial and antiinflammatory  
XX activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
XX antibodies that bind (I) are used in the manufacture of medicaments for  
XX the treatment or prevention of infection or disease caused by  
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX Nucleic acids encoding (I) are used to detect Streptococcus in a  
XX biological sample. (I) is used to determine whether a compound binds to  
XX (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
XX used as a vaccine or diagnostic composition. The disease caused by  
XX Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX acid encoding (I) may be used to recombinantly produce (I) and may be  
XX used in gene therapy. Antibodies to (I) are used for affinity  
XX chromatography, immunoassays, and distinguishing/identifying  
XX Streptococcus proteins

SQ Sequence 1731 BP; 519 A; 358 C; 389 G; 467 T; 0 U; 0 Other;  
Query Match 86.1%; Score 19.8; DB 6; Length 1731;  
Best Local Similarity 91.3%; Pred. No. 60;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 TGAGGAGGCGAGTGAAGCTTAG 23  
DB 177 TGAATAATGCAGTAGAAGCTTAG 199  
RESULT 29  
ACAS0572  
ID ACAS0572 standard; DNA; 1734 BP.  
XX AC ACAS0572;  
DT 19-JUN-2003 (first entry)  
XX Prokaryotic essential gene #32229.  
DE Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX Streptococcus pyogenes.  
OS WO20027183-A2.  
PN 03-OCT-2002.  
PD 21-MAR-2002; 2002WO-US009107.  
PP 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU46702.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX Claim 14; SEQ ID NO 38442; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 1734 BP; 519 A; 358 C; 389 G; 468 T; 0 U; 0 Other;

Query Match 86.1%; Score 19.8; DB 8; Length 1734;  
Best Local Similarity 91.3%; Pred. No. 60;  
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGGAGGCGAGTGAAGCTTAG 23  
DB 177 TGAATAATGCAGTAGAAGCTTAG 199

RESULT 30

ADD68806

ID ADD68806 standard; DNA; 1803 BP.

XX ADD68806;

DT 15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' *ptxI* DNA - SEQ ID 5.

DE beta-haemolytic Group A *Streptococcus*; GAS;

KW fluorescence resonance energy transfer; *PRET*; *ptxI*; *ds*; gene.

XX Streptococcus sp. 'group A'.

EH Key Location/Qualifiers

FT CDS 1..1803

FT /tag= a

FT /partial

FT /product= "Streptococcus sp. 'group A' *ptxI* protein - SEQ

FT ID 5"

FT /note= "No stop codon; this CDS contains translation

FT exceptions"

XX US6593093-B1.

XX 15-JUL-2003.

XX 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2003-828265/77.

XX P-PSDB; ADD68790.

PT Detecting presence/absence of group A *Streptococcus* in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.

XX Example 2; SEQ ID NO 5; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A *Streptococcus* (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridising step comprising contacting the



QY 2 GAGAGGCGAGTGAAGCTTAG 23  
DB 4526 GAGAGGCGAGTGAAGCTTAG 4505  
RESULT 33  
ADA14456/c  
ID ADA14456 standard; cDNA; 720 BP.  
AC ADA14456;  
XX  
XX 06-NOV-2003 (first entry)  
DT  
DE Mouse spermatogenesis related cDNA sequence SEQ ID NO:15.  
XX  
XX mouse; spermatogenesis; gene cluster; mutagenicity;  
KW reproductive toxicity; reproductive capacity; mutation;  
KW expression abnormality; human male sterility associated gene; scot-t;  
KW succinyl CoA:3-oxo acid CoA transferase; human male sterility; gene; ss.  
XX  
OS Mus musculus.  
XX  
XX WO2003068969-A1.  
XX  
XX 21-AUG-2003.  
XX  
XX 14-FEB-2003; 2003WO-JP001572.  
XX  
XX 14-FEB-2002; 2002JP-00036649.  
XX  
XX 27-DEC-2002; 2002JP-00381241.  
XX  
XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
XX  
XX Nishimune Y, Tanaka H, Nozaki M;  
XX WPI: 2003-671663/63.  
XX  
XX Mouse spermatogenesis gene cluster and human male sterility associated  
PT genes, useful for diagnosis of human male sterility and testing  
PT substances for reproductive toxicity.  
XX  
XX Claim 1; Page 70-71; 262pp; Japanese.  
XX  
XX The present invention describes a mouse spermatogenesis gene cluster  
CC containing 89 genes (see the cDNA sequences of ADA14442 to ADA14530).  
CC Also described: (1) a cDNA library containing cDNA encoding the gene  
CC cluster; (2) oligonucleotides of 10-99 bases containing partial sequences  
CC of genes of the cluster; (3) microarrays containing these  
CC oligonucleotides; (4) primer sets for PCR amplification of cDNA or  
CC genomic DNA for genes of the cluster; (5) polypeptides encoded by the  
CC genes in the cluster; (6) antibodies to these polypeptides; (7) a method  
CC for testing the mutagenicity and reproductive toxicity of a test  
CC substance, and assessment of the reproductive capacity of a test  
CC individual, by analysis of mutation and expression abnormalities of genes  
CC in the cluster; (8) polynucleotides which are mutations of the human male  
CC sterility associated gene scot-t (succinyl CoA:3-oxo acid CoA transferase  
CC gene) having one or more of the following specific mutations: T129C,  
CC T870G, C1071T, T1667C; (9) oligonucleotides containing partial sequences  
CC of human scot-t including one or more of the above mutations; (10) primer  
CC sets for PCR amplification of mRNA derived from the mutant scot-t gene;  
CC (11) polypeptides encoded by human scot-t gene and having one or more of  
CC the mutations Leu38Pro, Leu285Arg, Thr352Met; (12) polynucleotides which  
CC are mutations of the human male sterility associated gene protamine2,  
CC having C248T; (13) polypeptides encoded by this mutant protamine2 gene;  
CC (14) antibodies (including labelled antibodies) to these polypeptides;  
CC (15) a method for determining the presence or absence of these mutant  
CC polynucleotides in genomic DNA; (16) diagnosis of human male sterility  
CC using this method; (17) DNA probes containing sequences of these mutant  
CC polynucleotides; and (18) DNA chip containing sequences derived from  
CC these mutant polynucleotides. The methods of the present invention can be  
CC used in the diagnosis of human male sterility; testing the reproductive  
CC toxicity and mutagenicity of substances; and assessing the reproductive

CC capacity of individuals.  
XX  
XX Sequence 720 BP; 171 A; 200 C; 171 G; 178 T; 0 U; 0 Other;  
SQ  
Query Match 79.1%; Score 18.2; DB 9; Length 720;  
Best Local Similarity 87.0%; Pred. No. 2.8e+02;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCGAGTGAAGCTTAG 23  
DB 499 TGAGAAGGCGAGTGAAGCTTAG 477  
RESULT 34  
ADE56804  
ID ADE56804 standard; DNA; 15766 BP.  
XX  
XX ADE56804;  
XX  
XX 29-JAN-2004 (first entry)  
DT  
XX  
XX Rat gene J05677, SEQ ID NO 2659.  
DE  
XX  
XX Rat; ds; gene; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
XX Rattus norvegicus.  
OS  
XX  
XX WO2003016475-A2.  
XX  
XX 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
XX 14-AUG-2001; 2001US-0312147P.  
XX  
XX 01-NOV-2001; 2001US-0346382P.  
XX  
XX 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEHO) GEN HOSPITAL CORP.  
XX  
XX (FARB) BAYER AG.  
XX  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX WPI: 2003-268312/26.  
XX  
XX GENBANK; J05677.  
XX  
XX New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.  
XX  
XX The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction



```
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a rat DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 15766 BP; 3852 A; 4174 C; 3768 G; 3972 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 10; Length 15766;
Best Local Similarity 87.0%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
Dy 6272 TGTGAGGCGAGTAGAAGCATAG 6294
RESULT 35
AAV21209_01
Continuation (2 of 17) of AAV21209 from base 100001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209
WP Fragment Name Begin End
WP AAV21209_00 1 110000
WP AAV21209_01 100001 210000
WP AAV21209_02 200001 310000
WP AAV21209_03 300001 410000
WP AAV21209_04 400001 510000
WP AAV21209_05 500001 610000
WP AAV21209_06 600001 710000
WP AAV21209_07 700001 810000
WP AAV21209_08 800001 910000
WP AAV21209_09 900001 1010000
WP AAV21209_10 1000001 1110000
WP AAV21209_11 1100001 1210000
WP AAV21209_12 1200001 1310000
WP AAV21209_13 1300001 1410000
WP AAV21209_14 1400001 1510000
WP AAV21209_15 1500001 1610000
WP AAV21209_16 1600001 1664976
Query Match 79.1%; Score 18.2; DB 2; Length 110000;
Best Local Similarity 87.0%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
Dy 69531 TGAAAAGGCATTAAAAAGCTTAG 69553
RESULT 36
ABQ76673
ID ABQ76673 standard; DNA; 160820 BP.
XX
AC ABQ76673;
XX
DT 26-MAR-2003 (first entry)
XX
DE Androgen receptor signalling pathway-associated DNA AB043547.
XX
KW Androgen receptor; transactivation; modulator; Smad3; Smad4; Akt; TGF-B;
KW signal transduction pathway; transforming growth factor-B; phosphatase;
KW tensin; cytoskeletal; antiproliferative; cellular proliferation; cancer;
KW AB043547; ds.
XX
OS Synthetic.
XX
FN WO200282081-A2.
XX
PD 17-OCT-2002.
XX
PF 05-APR-2002; 2002WO-US011086.
XX
```

```
XX 06-APR-2001; 2001US-0282266P.
PR 13-MAR-2002; 2002US-0365060P.
XX
PA (UYRP ) UNIV ROCHESTER.
XX
PI Chang C;
XX
DR WPI; 2003-046871/04.
XX
PT Modulating androgen receptor activity, by administering a compound that
PT modulates receptor activity, inhibits receptor-signal transduction
PT pathway/receptor-coactivator interaction or changes amount or receptor.
XX
PS Disclosure; Page 181-225; 302pp; English.
XX
CC This invention describes a novel method for modulating androgen receptor
CC activity or androgen receptor-mediated transactivation activity in a
CC cell. The method involves administering a compound which causes
CC modulation of the androgen receptors activity and the inhibition of
CC interaction between the receptor and a protein involved in a signal
CC transduction pathway. The compound also inhibits the interaction between
CC the androgen receptor and a protein selected from Smad3, Smad4, Akt,
CC transforming growth factor (TGF)-B and phosphatase and tensin homologues
CC deleted on chromosome 10 (PTEN) or their fragments. The compounds of the
CC invention have cytostatic and antiproliferative activity. The obtained
CC composition is useful for treating any disease, where uncontrolled
CC proliferation or cellular proliferation occurs such as cancer, e.g.
CC prostate cancer. This sequence represents the androgen receptor
CC transactivation signalling pathway modulator AB043547 described in the
CC method of the invention
XX
SQ Sequence 160820 BP; 46593 A; 31350 C; 32677 G; 50200 T; 0 U; 0 Other;
Query Match 79.1%; Score 18.2; DB 8; Length 160820;
Best Local Similarity 87.0%; Pred. No. 4.7e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
Dy 55080 TGAGAAGGCGAGTAGGAAGCTTTG 55102
RESULT 37
ADCS9504
ID ADCS9504 standard; cDNA; 1423 BP.
XX
AC ADCS9504;
XX
DT 18-DEC-2003 (first entry)
XX
DE Polypeptide-respiratory chain dehydrogenase-11.55 cDNA.
XX
KW Polypeptide-respiratory chain dehydrogenase-11.55; gene; ss; cancer;
KW malignant tumour; haemopathy; human immunodeficiency virus infection;
KW HIV; immunological disease; inflammation; antagonist.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FH CDS 225..542
FT /*tag= a
FT /product= "Polypeptide-respiratory chain dehydrogenase-
FT 11.55"
FT /note= "This CDS is specifically claimed in claim 6"
XX
FN CNL1381565-A.
XX
PD 27-NOV-2002.
XX
PF 18-APR-2001; 2001CN-00112624.
XX
PR 18-APR-2001; 2001CN-00112624.
```



PT developing a diagnostic marker or medicines for regulating their  
PT expression and activity, or as a target of gene therapy.  
XX  
PS Claim 1; SEQ ID NO 1480; 305pp; English.  
XX  
CC The invention relates to a novel human polynucleotide and the encoded  
CC polypeptide. A polynucleotide of the invention may have a use in gene  
CC therapy. An oligonucleotide of the invention ADM06202-ADM06773 is useful  
CC as a primer for synthesizing the polynucleotide or as a probe for  
CC detecting the polynucleotide. The polynucleotides ADM01316-ADM03758 are  
CC useful in gene therapy, for developing a diagnostic marker or medicines  
CC for regulating their expression and activity, or as a target of gene  
CC therapy. The proteins ADM03759-ADM06201 encoded by the polynucleotides  
CC are useful as pharmaceutical agents. The present sequence represents a  
CC cDNA sequence of the invention.  
XX  
SQ Sequence 2533 BP; 590 A; 614 C; 707 G; 622 T; 0 U; 0 Other;  
Query Match 77.4%; Score 17.8; DB 11; Length 2533;  
Best Local Similarity 90.5%; Pred. No. 4.8e+02; Length 2533;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TGAGAAGGCAGTAGAAGCTT 21  
Db 702 TGAGAAGGCAGTAGAAGCTT 682  
RESULT 40  
ID ABL13110  
XX ABL13110 standard; cDNA; 6805 BP.  
XX  
AC ABL13110;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33812.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; sb.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PsDB; ABB69007.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Claim 1; SEQ ID NO 33812; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX Sequence 6805 BP; 2201 A; 1410 C; 1307 G; 1887 T; 0 U; 0 Other;  
SQ  
Query Match 77.4%; Score 17.8; DB 4; Length 6805;  
Best Local Similarity 90.5%; Pred. No. 5.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 TGAGAAGGCAGTAGAAGCTT 21  
Db 2650 TGAGAAGGCAGTAGAAGCTT 2670  
RESULT 41  
ID ABA15166 standard; DNA; 9722 BP.  
XX  
AC ABA15166;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 7497.  
XX  
KW Human; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200159063-A2.  
XX  
PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001334.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0188874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225575P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.



ID ABAL15164 standard; DNA; 14855 BP.  
XX ABAL15164;  
AC  
XX  
XX  
DT 23-JAN-2002 (first entry)  
XX  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 7495.  
XX  
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
KW antiparkinsonian; antisclerotic; antianemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200159063-A2.  
PN  
XX  
PD 16-AUG-2001.  
XX  
XX  
PF 17-JAN-2001; 2001WO-US001334.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0191213P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216800P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0231415P.  
PR 08-SEP-2000; 2000US-0231416P.  
PR 08-SEP-2000; 2000US-0231417P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231988P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241836P.  
PR 20-OCT-2000; 2000US-0242211P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.

```
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250391P.
PR 01-DEC-2000; 2000US-0251160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
XX Disclosure; SEQ ID NO 7495; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABA21534) and proteins
CC (ABBI4678-ABBI4801) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, diabetes mellitus, Crohn's disease,
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 14855 BP; 3335 A; 3554 C; 4013 G; 3953 T; 0 U; 0 Other;
SQ
Query Match 77.4%; Score 17.8; DB 5; Length 14855;
Best Local Similarity 90.5%; Pred. No. 5.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGCTAGAAAGCTT 21
Db 8160 TGAGAGGCAGCTAGAAAGCTT 8180

RESULT 43
ABA93286
ID ABA93286 standard; DNA; 22028 BP.
XX
XX ABA93286;
AC
XX
XX 22-APR-2002 (first entry)
XX
XX Human acetyl-Coenzyme A acyltransferase 1 gene SEQ ID NO:1.
DE
XX Human; acetyl-Coenzyme A acyltransferase; ACAA1; chromosome 3p23-p22;
KW peroxisomal 3-oxoacyl-Coenzyme A thiolase; SNP; genotype; haplotype;
KW single nucleotide polymorphism; polymorphic variant; enzyme; gene; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH
```

```
FT variation
FT /replace(3546,A)
FT /*tag= a
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(3537,A)
FT /*tag= b
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(4033,T)
FT /*tag= c
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(4157,T)
FT /*tag= d
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(4176,A)
FT /*tag= e
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(4276,C)
FT /*tag= f
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(9110,G)
FT /*tag= g
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(9182,A)
FT /*tag= h
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(9600,G)
FT /*tag= i
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(11702,T)
FT /*tag= j
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(11756,C)
FT /*tag= k
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(13390,T)
FT /*tag= l
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(14441,G)
FT /*tag= m
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(14872,C)
FT /*tag= n
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(14931,C)
FT /*tag= o
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(15471,C)
FT /*tag= p
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(15486,G)
FT /*tag= q
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(18004,T)
FT /*tag= r
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
FT variation
FT /replace(18192,T)
FT /*tag= s
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT
XX WO200187903-A2.
XX
XX 22-NOV-2001.
XX
XX 03-MAY-2001; 2001WO-US014330.
XX
XX 18-MAY-2000; 2000US-0205022P.
XX (GENA-) GENAISANCE PHARM INC.
XX (DUDA/) DUDA A E.
XX Chew A, Koshy B;
XX WPI; 2002-164134/21.
XX
```

PT Isolated polynucleotide, comprising a polymorphic variant of the acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase) gene useful for providing haplotype information and in therapy for treating related disorders.

XX Claim 19; Fig 1A-I; 93pp; English.

XX The present invention describes a polypeptide (I) which is a polymorphic variant (PV) of the acetyl-Coenzyme A acyltransferase (peroxisomal 3-oxoacyl-Coenzyme A thiolase) ACAAI protein (ABR05516). ACAAI is located on chromosome 3p23-p22. (I) can be encoded by ABA93286 (or ABA93288) CC where the sequence comprises one of the haplotypes shown in Table 4 or CC one of the haplotype pairs shown in Table 3, where Tables 3 and 4 are CC given in the specification. The polynucleotide encoding ACAAI can be used CC for providing haplotype and genotype information of an individual. CC Furthermore, the polynucleotide is useful for the treatment of disorders CC related to its abnormal expression or function. The present sequence CC represents the human ACAAI gene, featuring all the single nucleotide CC polymorphisms (SNP) possible, from the present invention

XX SQ Sequence 22028 BP; 4811 A; 5599 C; 6137 G; 5481 T; 0 U; 0 Other;

Query Match 77.4%; Score 17.8; DB 6; Length 22028;  
Best Local Similarity 90.5%; Pred. No. 5.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 12094 TGAGAAGGCAGTGGGAGCTT 12114

RESULT 44  
ABA93288  
ID ABA93288 standard; DNA; 22028 BP.

AC ABA93288;

DT 22-APR-2002 (first entry)

DE Human acetyl-Coenzyme A acyltransferase 1 gene SEQ ID NO:99.

XX Human; acetyl-Coenzyme A acyltransferase; ACAAI; chromosome 3p23-p22;  
KW peroxisomal 3-oxoacyl-Coenzyme A thiolase; SNP; genotype; haplotype;  
XX single nucleotide polymorphism; polymorphic variant; enzyme; gene; ds.

OS Homo sapiens.

PN WO200187903-A2.

PD 22-NOV-2001.

PF 03-MAY-2001; 2001WO-US014330.

PR 18-MAY-2000; 2000US-0205022P.

XX (GENA-) GENAISSANCE PHARM INC.  
PA (DUDA/) DUDA A E.

PI Chew A, Koshy B;

XX WPI; 2002-164134/21.

XX Isolated polynucleotide, comprising a polymorphic variant of the acetyl-Coenzyme A acyltransferase 1 (peroxisomal 3-oxoacyl-Coenzyme A thiolase) gene useful for providing haplotype information and in therapy for treating related disorders.

PS Disclosure; Page 85-93; 93pp; English.

XX The present invention describes a polypeptide (I) which is a polymorphic variant (PV) of the acetyl-Coenzyme A acyltransferase (peroxisomal 3-oxoacyl-Coenzyme A thiolase) ACAAI protein (ABR05516). ACAAI is located CC on chromosome 3p23-p22. (I) can be encoded by ABA93286 (or ABA93288)

CC where the sequence comprises one of the haplotypes shown in Table 4 or CC one of the haplotype pairs shown in Table 3, where Tables 3 and 4 are CC given in the specification. The polynucleotide encoding ACAAI can be used CC for providing haplotype and genotype information of an individual. CC Furthermore, the polynucleotide is useful for the treatment of disorders CC related to its abnormal expression or function. The present sequence CC represents the human ACAAI gene, where all the single nucleotide CC polymorphisms (SNP) possible have been given in the sequence using the CC standard ambiguity codes

XX SQ Sequence 22028 BP; 4808 A; 5592 C; 6133 G; 5476 T; 0 U; 19 Other;

Query Match 77.4%; Score 17.8; DB 6; Length 22028;  
Best Local Similarity 90.5%; Pred. No. 5.9e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
Db 12094 TGAGAAGGCAGTGGGAGCTT 12114

RESULT 45  
ABD33148/c

ID ABD33148 standard; DNA; 104399 BP.

AC ABD33148;

DT 18-NOV-2004 (first entry)

DE Murine cancer-associated (CA) gene MD07-019.

XX Mouse; cancer-associated protein; CAP; cancer-associated gene; CA; gene;  
KW ds; cancer; cytostatic.

OS Mus musculus.

PN WO2004058146-A2.

PD 15-JUL-2004.

PF 15-DEC-2003; 2003WO-US040081.

PR 17-DEC-2002; 2002US-00322281.

XX (SAGR-) SAGRES DISCOVERY INC.

PI Morris DW, Malandro MS;

XX WPI; 2004-499109/47.

XX Novel human cancer associated protein encoded within open reading frame  
PT of cancer associated gene, useful as targets for diagnosing cancer.

PS Disclosure; SEQ ID NO 123; 182pp; English.

XX The invention relates to cancer-associated proteins (CAP) and the cancer-associated (CA) nucleic acids encoding them. The invention also relates CC to a method for treating cancers involving administering to a patient an CC inhibitor of CAP, and a method of screening for anticancer activity in a CC potential drug involving providing a cell that expresses a CA gene, CC contacting a tissue sample derived from a cancer cell with an anticancer CC drug candidate and monitoring the effect of the anticancer drug candidate CC on expression of the CA gene. The CAP proteins are useful for detecting CC cancer associated with expression of a CAP protein in a test cell sample CC and for screening for a bioactive agent capable of modulating the CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing CC cancer, involving determining the expression of a CA nucleic acid in a CC tissue. This sequence represents a murine CA gene of the invention. Note: CC The sequence data for this patent did not form part of the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 104399 BP; 25540 A; 20083 C; 20918 G; 29473 T; 0 U; 8385 Other;



```
Query Match      77.4%; Score 17.8; DB 13; Length 104399;
Best Local Similarity 90.5%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GAGAGGCAGTAGAAAGCTTAG 22
DB      35008 GAGAGGCAGTAGAAAGACTA 34988

RESULT 46
AAV21209_12/c
Continuation (13 of 17) of AAV21209 from base 1200001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209
WP Fragment Name      Begin      End
WP AAV21209_00        1      110000
WP AAV21209_01       100001    210000
WP AAV21209_02       200001    310000
WP AAV21209_03       300001    410000
WP AAV21209_04       400001    510000
WP AAV21209_05       500001    610000
WP AAV21209_06       600001    710000
WP AAV21209_07       700001    810000
WP AAV21209_08       800001    910000
WP AAV21209_09       900001   1010000
WP AAV21209_10      1000001   1110000
WP AAV21209_11      1100001   1210000
WP AAV21209_12      1200001   1310000
WP AAV21209_13      1300001   1410000
WP AAV21209_14      1400001   1510000
WP AAV21209_15      1500001   1610000
WP AAV21209_16      1600001   1664976

Query Match      77.4%; Score 17.8; DB 2; Length 110000;
Best Local Similarity 90.5%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTTAG 23
DB      105980 AGAAGGCAGTAGAAAGCTTAG 105960

RESULT 47
AAV21209_13/c
Continuation (14 of 17) of AAV21209 from base 1300001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209
WP Fragment Name      Begin      End
WP AAV21209_00        1      110000
WP AAV21209_01       100001    210000
WP AAV21209_02       200001    310000
WP AAV21209_03       300001    410000
WP AAV21209_04       400001    510000
WP AAV21209_05       500001    610000
WP AAV21209_06       600001    710000
WP AAV21209_07       700001    810000
WP AAV21209_08       800001    910000
WP AAV21209_09       900001   1010000
WP AAV21209_10      1000001   1110000
WP AAV21209_11      1100001   1210000
WP AAV21209_12      1200001   1310000
WP AAV21209_13      1300001   1410000
WP AAV21209_14      1400001   1510000
WP AAV21209_15      1500001   1610000
WP AAV21209_16      1600001   1664976

Query Match      77.4%; Score 17.8; DB 2; Length 110000;
Best Local Similarity 90.5%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTTAG 23
DB      5980 AGAAGGCAGTAGAAAGCTTAG 5960

RESULT 48
ADD68802
ID ADD68802 standard; DNA; 21 BP.
XX
AC ADD68802;
XX
DT 15-JAN-2004 (first entry)
XX
DE PCR primer ptsU used to amplify Group A Streptococcus ptsI sequences.
XX
KW beta-haemolytic Group A Streptococcus; GAS;
KW fluorescence resonance energy transfer; FRET; ptsU; PCR; primer; ptsI;
KW ss.
XX
OS Streptococcus sp. 'group A'.
XX
PN US6593093-B1.
XX
PD 15-JUL-2003.
XX
PF 20-FEB-2002; 2002US-00081923.
XX
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
DR WPI; 2003-828265/77.
XX
PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
PS Claim 2; SEQ ID NO 1; 35pp; English.
XX
CC The invention relates to a novel method for detecting the presence or
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
CC isolated from an individual by performing more than one cycling step. The
CC method involves an amplifying step comprising contacting the sample with
CC GAS primers, a subsequent hybridizing step comprising contacting the
CC amplified product with fluorescent-labelled probes and finally, detecting
CC fluorescence resonance energy transfer (FRET) where the presence or
CC absence of FRET indicates the presence or absence of GAS in the sample.
CC The method of the invention may be useful for detecting the presence or
CC absence of GAS in a biological sample from an individual. The current
CC sequence is that of the ptsU PCR primer of the invention which was used
CC to amplify GAS ptsI sequences.
XX
SQ Sequence 21 BP; 9 A; 2 C; 6 G; 4 T; 0 U; 0 Other;

Query Match      75.7%; Score 17.4; DB 10; Length 21;
Best Local Similarity 94.7%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      5 AGGCAGTAGAAAGCTTAG 23
DB      2 AATGCAGTAGAAAGCTTAG 20

RESULT 49
ADJ61989
ID ADJ61989 standard; DNA; 21 BP.
XX
AC ADJ61989;
XX
DT 06-MAY-2004 (first entry)
XX
DE ptsI gene PCR primer, ptsU.
XX
KW Group A Streptococcus; GAS; PCR; primer; ss.
```



CC compounds or the progression of these toxic effects by determining the  
CC Global changes in gene expression in tissues or cells exposed to the  
CC toxin and comparing these to gene expression in unexposed tissues or  
CC cells. Also included are methods of predicting at least one toxic effect  
CC of a compound or progression of a toxic effect, preferably the  
CC hepatotoxicity of a compound, comprising detecting the level of  
CC expression in a tissue or cell sample exposed to the compound of two or  
CC more genes listed in the specification, where differential expression of  
CC the genes is indicative of at least one toxic effect or progression. The  
CC method can also be used to identify an agent which modulates the toxic  
CC response and predict cellular pathways that a compound modulates in a  
CC cell. The methods utilize a set of at least two probes (on a solid  
CC support in kit form), where each of the probes comprises a sequence that  
CC specifically hybridises to a gene listed in the specification, a computer  
CC system comprising a database containing information identifying the  
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
CC set of genes comprising at least two genes listed in the specification,  
CC and a user interface to view the information used to present information  
CC identifying the expression level in a tissue or cell of at least one gene  
CC listed in the specification. The method is useful for elucidating global  
CC changes in gene expression and for identifying toxicity markers in  
CC tissues or cell exposed to a known toxin. The genes may be used as  
CC toxicity markers in drug screening and toxicity assays. The genes and  
CC gene expression information may be used as diagnostic markers for the  
CC prediction or identification of the physiological state of tissue or cell  
CC sample that has been exposed to a compound or agent. Hepatotoxicity is  
CC characterised by centrilobular necrosis and steatosis. The present  
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
CC which is differentially expressed in response to a hepatotoxic agent  
XX  
SQ Sequence 322 BP; 96 A; 73 C; 74 G; 79 T; 0 U; 0 Other;

Query Match 75.7%; Score 17.4; DB 6; Length 322;  
Best Local Similarity 94.7%; Pred. No. 6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TGAGAGGCGAGTGAAGC 19  
DB 252 TGAGAGGCGAGTGAAGC 234

RESULT 52  
ADZ62746  
ID ADZ62746 standard; DNA; 2151 BP.  
XX  
AC ADZ62746;  
XX  
DT 30-JUN-2005 (first entry)  
XX  
DE Murine Cdc45i gene, SEQ ID 2470.  
XX  
XX Drug screening; Peroxisome Proliferator-Activated Receptor alpha;  
KW PPAR alpha; gene; ds.  
XX  
XX Mus musculus.  
XX  
XX US2005084872-A1.  
XX  
XX 21-APR-2005.  
XX  
XX 23-JAN-2004; 2004US-00764420.  
XX  
XX 24-MAY-2003; 2003US-0442797P.  
XX 30-MAY-2003; 2003US-0474413P.  
XX  
XX (LUMP/) LUM P Y.  
XX (TANY/) TAN Y.  
XX (DAIH/) DAI H.  
XX (MUIS/) MUISE E S.  
XX (BERG/) BERGER J P.  
XX (THOM/) THOMPSON J R.  
XX  
XX Lum PY, Tan Y, Dai H, Muise ES, Berger JP, Thompson JR;

XX WPI; 2005-313963/32.  
DR REFSEQ; NM\_009862.  
XX  
XX Determining whether agent has defined biological activity, by making  
PT comparison e.g. comparing efficacy value, toxicity value and classifier  
PT value of agent to reference values, and using results to determine  
PT whether agent has activity.  
XX  
XX Example 4; SEQ ID NO 2470; 51pp; English.  
XX  
XX The present invention relates to a methods for screening biologically  
CC active agents, such as candidate drugs, to determine whether the agent  
CC possesses a defined biological activity. The methods involve making one  
CC or more comparisons chosen from comparing efficacy value of agent to  
CC reference efficacy value, comparing a toxicity value of the agent to  
CC reference toxicity value, comparing a classifier value of the agent to  
CC reference classifier value, and using the comparison result(s) to  
CC determine whether the agent possesses the defined biological activity.  
CC Also claimed is a population of oligonucleotide probes (1) specific for  
CC measuring the expression levels of members of a classifier population of  
CC genes or a toxicity-related population of genes. (1) are useful for  
CC measuring the expression levels of genes that are useful for identifying  
CC agonists or partial agonists of Peroxisome Proliferator-Activated  
CC Receptor (PPAR) gamma. ADZ63072-ADZ63959 are oligonucleotide probes which  
CC are useful in the method of the invention to measure the expression  
CC pattern of mouse genes ADZ62273-ADZ63071 that yield an expression pattern  
CC that correlates with the stimulation of PPARalpha receptors by an agent.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?DocID=20050084872.  
XX  
XX Sequence 2151 BP; 559 A; 499 C; 551 G; 542 T; 0 U; 0 Other;

Query Match 75.7%; Score 17.4; DB 14; Length 2151;  
Best Local Similarity 94.7%; Pred. No. 7.2e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TGAGAGGCGAGTGAAGC 19  
DB 1891 TGAGAGGCGAGTGAAGC 1909

RESULT 53  
AAK65192/C  
ID AAK65192 standard; DNA; 7613 BP.  
XX  
AC AAK65192;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:20004.  
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytotstatic; gene therapy; vaccine; metastasis; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
PR

PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0215647P.  
PR 07-JUL-2000; 2000US-0215680P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0228279P.  
PR 22-AUG-2000; 2000US-0228681P.  
PR 22-AUG-2000; 2000US-0228688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239936P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249266P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-02559678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
PS Disclosure; SEQ ID NO 20004; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to RAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX  
SQ Sequence 7613 BP; 1434 A; 2458 C; 2149 G; 1572 T; 0 U; 0 Other;

Query Match 75.7%; Score 17.4; DB 4; Length 7613;  
Best Local Similarity 94.7%; Pred. No. 8.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTT 21  
|||||  
DB 3366 AGAAGGCAGTAGAAGCTT 3348

RESULT 54  
ABL28640  
ID ABL28640 standard; DNA; 24789 BP.  
XX  
XX ABL28640;  
AC  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37393.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
XX  
PD 27-SEP-2001.  
XX  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX  
XX WPI; 2001-656860/75.  
XX  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.

XX  
XX  
XX Claim 1; SEQ ID NO 37393; 21pp + Sequence Listing; English.

XX  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pcr\_sequences

XX  
XX  
SQ Sequence 24789 BP; 7654 A; 5448 C; 6281 G; 5406 T; 0 U; 0 Other;  
Query Match 75.7%; Score 17.4; DB 4; Length 24789;  
Best Local Similarity 94.7%; Pred. No. 9e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TGAGAAGGCAGTAGAAGC 19  
|||||  
DB 13733 TGAGAAGGCAGTAGAAGC 13751

RESULT 55  
ABK62529  
ID ABK62529 standard; cDNA; 519 BP.  
XX  
XX  
XX ABK62529;  
AC  
XX  
DT 18-JUN-2002 (first entry)  
XX  
DE Rat sequence differentially expressed in response to a hepatotoxin #436.

XX  
XX  
XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;  
KW differential expression; centrilobular necrosis; steatosis.  
XX  
XX Rattus norvegicus.  
OS  
XX  
XX WO200210453-A2.  
XX  
XX 07-FEB-2002.  
XX  
XX 30-JUL-2001; 2001WO-US023872.  
XX  
XX 31-JUL-2000; 2000US-0222040P.  
PR 02-NOV-2000; 2000US-0244880P.  
PR 11-MAY-2001; 2001US-0290029P.  
PR 15-MAY-2001; 2001US-0290645P.  
PR 22-MAY-2001; 2001US-0292336P.  
PR 06-JUN-2001; 2001US-0295798P.  
PR 19-JUN-2001; 2001US-0297457P.  
PR 09-JUL-2001; 2001US-0303459P.  
XX  
XX (GENE-) GENE LOGIC INC.  
XX  
XX Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;  
PI  
XX  
XX WPI; 2002-241625/29.

XX  
XX  
XX Predicting toxic effects of compounds or the progression of these toxic  
PT effects by determining the changes in gene expression in tissues or cells  
PT exposed to the toxin and comparing these to gene expression in unexposed  
PT tissues or cells.

XX  
XX  
XX Claim 1; SEQ ID NO 436; 239pp; English.

XX  
XX  
XX The invention relates to methods for predicting toxic effects of  
CC compounds or the progression of these toxic effects by determining the  
CC global changes in gene expression in tissues or cells exposed to the  
CC toxin and comparing these to gene expression in unexposed tissues or  
CC cells. Also included are methods of predicting at least one toxic effect  
CC of a compound or progression of a toxic effect, preferably the  
CC hepatotoxicity of a compound, comprising detecting the level of  
CC expression in a tissue or cell sample exposed to the compound of two or  
CC more genes listed in the specification, where differential expression of  
CC the genes is indicative of at least one toxic effect or progression. The  
CC method can also be used to identify an agent which modulates the toxic  
CC response and predict cellular pathways that a compound modulates in a  
CC cell. The methods utilise a set of at least two probes (on a solid  
CC support in kit form), where each of the probes comprises a sequence that  
CC specifically hybridises to a gene listed in the specification, a computer  
CC system comprising a database containing information identifying the  
CC expression level in a tissue or cell sample exposed to a hepatotoxin of a  
CC set of genes comprising at least two genes listed in the specification,  
CC and a user interface to view the information used to present information,  
CC identifying the expression level in a tissue or cell of at least one gene  
CC listed in the specification. The method is useful for elucidating global  
CC changes in gene expression and for identifying toxicity markers in

CC tissues or cell exposed to a known toxin. The genes may be used as  
CC toxicity markers in drug screening and toxicity assays. The genes and  
CC gene expression information may be used as diagnostic markers for the  
CC prediction or identification of the physiological state of tissue or cell  
CC sample that has been exposed to a compound or agent. Hepatotoxicity is  
CC characterized by centrilobular necrosis and steatosis. The present  
CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene  
CC which is differentially expressed in response to a hepatotoxic agent  
XX  
XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 6; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGGCAGTAGAAAGCTTAG 200

RESULT 56  
ADBS6078  
ID ADBS6078 standard; DNA; 519 BP.

XX ADBS6078;

DT 04-DEC-2003 (first entry)

DE Toxicity-related gene, SEQ ID 1104.

KW Toxic; toxin; gene expression profile; hepatotoxicity; liver;  
KW drug screening; toxicity assay; ds.

XX Unidentified.

OS WO2003064624-A2.

PN 07-AUG-2003.

PD 31-JAN-2003; 2003WO-US003194.

PR 31-JAN-2002; 2002US-00060087.

PR 15-MAR-2002; 2002US-0364045P.

PR 15-MAR-2002; 2002US-0364055P.

PR 30-DEC-2002; 2002US-0436643P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Elashoff M;

XX WPI; 2003-689530/65.

XX Predicting a toxic effect of a compound, useful in identifying toxicity  
XX markers in liver tissues or cells for drug screening and toxicity assays,  
XX comprises preparing gene expression profile of tissue or cells exposed to  
XX the compound.

XX Claim 1; SEQ ID NO 1104; 1156pp; English.

XX The present invention relates to a method for predicting a toxic effect  
XX of a compound. The method comprises preparing a gene expression profile  
XX of a tissue or cell sample exposed to the compound, and comparing the  
XX gene expression profile to a database comprising SEQ ID 1-4925, where  
XX differential expression of the gene indicates at least one toxic effect.  
XX The method is useful for predicting at least one toxic effect of a  
XX compound, predicting hepatotoxicity or the progression of a toxic effect  
XX of a compound, identifying an agent that modulates the onset or  
XX progression of a toxic response, predicting the cellular pathways that a  
XX compound modulates in a cell, and identifying an agent that modulates at  
XX least one activity of a protein. The method and compositions of the  
XX present invention using a database of genes having liver toxin-induced  
XX differential expression, are useful in identifying toxicity markers in  
XX liver tissues or cells for drug screening and toxicity assays. Note: The

CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;

Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGCTTAG 23  
Db 179 GCGAAGGCAGTAGAAAGCTTAG 200

RESULT 57  
ADBS0614  
ID ADBS0614 standard; DNA; 519 BP.

XX ADBS0614;

DT 04-DEC-2003 (first entry)

DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:1156.

KW toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;  
KW toxicity marker; toxicity progression; drug screening;

XX primary rat hepatocyte toxicity modelling; gene; ds.

OS Rattus norvegicus.

PN WO2003065993-A2.

XX 14-AUG-2003.

XX 04-FEB-2003; 2003WO-US003482.

XX 04-FEB-2002; 2002US-0353171P.

XX 13-MAR-2002; 2002US-0363534P.

XX 08-APR-2002; 2002US-0370248P.

XX 10-APR-2002; 2002US-0371134P.

XX 10-APR-2002; 2002US-0371135P.

XX 11-APR-2002; 2002US-0371150P.

XX 11-APR-2002; 2002US-0371413P.

XX 19-APR-2002; 2002US-0373601P.

XX 19-APR-2002; 2002US-0373602P.

XX 22-APR-2002; 2002US-0374139P.

XX 08-MAY-2002; 2002US-0378370P.

XX 09-MAY-2002; 2002US-0378652P.

XX 09-MAY-2002; 2002US-0378653P.

XX 09-JUL-2002; 2002US-0394230P.

XX 09-JUL-2002; 2002US-0394253P.

XX 04-SEP-2002; 2002US-0407688P.

XX 28-JAN-2003; 2003US-0442900P.

XX (GENE-) GENE LOGIC INC.

XX Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;

XX Elashoff M;

XX WPI; 2003-731472/69.

XX Determining if a compound induces a toxic effect on a tissue or cell, for  
XX identifying hepatotoxic compounds, comprises comparing a gene expression  
XX profile of a tissue or cell sample to a database of Tox mean and non-Tox  
XX mean values.

XX Claim 44; SEQ ID NO 1156; 874pp; English.

XX The present invention describes a method for determining whether a  
XX compound induces a toxic effect on a tissue or cell. The method comprises  
XX preparing a gene expression profile of a tissue or cell sample exposed to

CC the compound, and comparing the gene expression profile to a database  
CC comprising data or information on the Tox mean and non-Tox mean value.  
CC The method is useful for predicting or identifying at least one toxic  
CC effect, particularly hepatotoxicity, of a test or unknown compound. The  
CC genes listed in the specification are useful as diagnostic or toxicity  
CC markers for the prediction or identification of the physiological state  
CC of tissue or cell sample that has been exposed to a compound, or to  
CC identify or predict the toxic effects of a compound or an agent. These  
CC may also be used as markers for monitoring toxicity progression or for  
CC drug screening. The present sequence represents a primary rat hepatocyte  
CC toxicity modelling related gene sequence from the present invention.  
XX  
SQ Sequence 519 BP; 156 A; 113 C; 114 G; 135 T; 0 U; 1 Other;  
  
Query Match 74.8%; Score 17.2; DB 10; Length 519;  
Best Local Similarity 86.4%; Pred. No. 7.7e-02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 2 GAGAAGGCAGTAGAAGCTTAG 23  
DB 179 GGGAAAGCAGTAGAAGCTTAG 200  
  
RESULT 58  
ABL90230/c  
ID ABL90230 standard; cDNA; 670 BP.  
XX  
AC ABL90230;  
DT 24-MAY-2002 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 792.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200190304-A2.  
XX  
PD 29-NOV-2001.  
XX  
PF 18-MAY-2001; 2001WO-US016450.  
XX  
PR 19-MAY-2000; 2000US-0205515P.  
XX  
PS (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Birse CE, Rosen CA;  
XX  
DR WPI: 2002-122018/16.  
DR P-PSDB; ABB98821.  
XX  
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders.  
XX  
PS Claim 4; SEQ ID NO 792; 2081pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABL9449-ABL90853) and proteins  
CC (ABB9040-ABB90444) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC

CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 670 BP; 182 A; 133 C; 115 G; 229 T; 0 U; 11 Other;

Query Match 74.8%; Score 17.2; DB 6; Length 670;  
Best Local Similarity 86.4%; Pred. No. 7.9e-02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23  
DB 549 GAGAAGGCAGTAGAAGGAGAG 528

RESULT 59  
AAX97673/c  
ID AAX97673 standard; DNA; 886 BP.  
XX  
AC AAX97673;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE Extended human secreted protein coding sequence, SEQ ID NO. 137.

XX  
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;  
KW cellular differentiation; immune system regulator; anti-inflammatory;  
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;  
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;  
KW genetic disease; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9911236-A2.  
XX  
PD 24-JUN-1999.  
XX  
PF 17-DEC-1998; 98WO-IB002122.  
XX  
PR 17-DEC-1997; 97US-0069957P.  
PR 09-FEB-1998; 98US-0074121P.  
PR 13-APR-1998; 98US-0081563P.  
PR 10-AUG-1998; 98US-0096116P.  
XX  
PA (GEST ) GENSET.  
XX  
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;  
XX  
DR WPI: 1999-385906/32.  
DR P-PSDB; AAY35989.  
XX  
PT New isolated human secreted proteins.  
XX  
PS Claim 1; Page 265-266; 516pp; English.

XX  
CC This sequence represents an extended human secreted protein coding  
CC sequence of the invention. The secreted proteins can be used in treating  
CC or controlling a variety of human conditions. The secreted proteins may  
CC act as cytokines or may affect cellular proliferation or differentiation  
CC or may act as immune system regulators, haematopoiesis regulators, tissue  
CC growth regulators, regulators of reproductive hormones or cell movement  
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or  
CC tumour inhibition activity. The DNAs can be used in forensic procedures  
CC to identify individuals or in diagnostic procedures to identify  
CC individuals having genetic diseases resulting from abnormal expression of  
CC the genes corresponding to the extended cDNAs. They are also useful for  
CC constructing a high resolution map of the human chromosomes. They can  
CC also be used for gene therapy to control or treat genetic diseases  
XX





GenCore version 5.1.7  
Copyright (C) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 20:43:21 ; Search time 196,315 Seconds  
(without alignments)  
712.930 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcagtgtggtttatcttcc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 60 summaries

Database :

- N Geneseq\_21.\*
- 1: Geneseqn1980s.\*
- 2: Geneseqn1990s.\*
- 3: Geneseqn2000s.\*
- 4: Geneseqn2001as.\*
- 5: Geneseqn2001bs.\*
- 6: Geneseqn2002as.\*
- 7: Geneseqn2002bs.\*
- 8: Geneseqn2003as.\*
- 9: Geneseqn2003bs.\*
- 10: Geneseqn2003cs.\*
- 11: Geneseqn2003ds.\*
- 12: Geneseqn2004as.\*
- 13: Geneseqn2004bs.\*
- 14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	10	ADD68803
2	21	100.0	21	12	ADJ61990
3	21	100.0	21	14	ADZ75862
C 4	21	100.0	1323	10	ADD68796
C 5	21	100.0	1323	12	ADJ62000
C 6	21	100.0	1378	10	ADD68798
C 7	21	100.0	1378	12	ADJ62002
C 8	21	100.0	1379	10	ADD68800
C 9	21	100.0	1379	10	ADD68797
C 10	21	100.0	1379	10	ADD68791
C 11	21	100.0	1379	12	ADJ62001
C 12	21	100.0	1379	12	ADJ62004
C 13	21	100.0	1379	12	ADJ61995
C 14	21	100.0	1384	10	ADD68792
C 15	21	100.0	1384	10	ADD68794
C 16	21	100.0	1384	12	ADJ61998
C 17	21	100.0	1384	12	ADJ61996
C 18	21	100.0	1385	10	ADD68793
C 19	21	100.0	1385	12	ADJ61997

C 20	21	100.0	1390	10	ADD68795
C 21	21	100.0	1390	12	ADJ61999
C 22	21	100.0	1393	10	ADD68799
C 23	21	100.0	1393	12	ADJ62003
C 24	21	100.0	1447	10	ADD68801
C 25	21	100.0	1447	12	ADJ62005
C 26	21	100.0	1731	6	ABN68277
C 27	21	100.0	1731	6	ABN68276
C 28	21	100.0	1734	8	ACA50572
C 29	21	100.0	1734	13	ADV85417
C 30	21	100.0	1803	10	ADD68806
C 31	21	100.0	1803	12	ADJ61993
C 32	21	100.0	29072	13	ADV87723
C 33	21	100.0	29072	13	ADV78976
C 34	21	100.0	110000	6	ABN71527_07
C 35	21	100.0	110000	13	ADV81204_08
C 36	19.4	92.4	402	10	ABX06888
C 37	19.4	92.4	962	2	AAV65250
C 38	19.4	92.4	1140	13	ADK44676
C 39	19.4	92.4	1287	13	ADR92834
C 40	19.4	92.4	1287	14	AEA56704
C 41	19.4	92.4	1734	4	AA555546
C 42	19.4	92.4	1734	8	ACA49934
C 43	19.4	92.4	8195	2	AAV52227
C 44	19.4	92.4	110000	10	ABS56454_09
C 45	19.4	92.4	110000	10	ABS56454_10
C 46	18.4	87.6	553	12	ACH68181
C 47	17.8	84.8	1830	9	ADB11789
C 48	17.8	84.8	1830	9	ADB11785
C 49	17.8	84.8	1830	9	ADB11787
C 50	17.8	84.8	1830	9	ADB11783
C 51	17.8	84.8	2914	7	ADR32245
C 52	17.8	84.8	2914	10	ADA53346
C 53	17.8	84.8	110000	9	ADB12064_13
C 54	16.8	80.0	412	6	ABK74374
C 55	16.8	80.0	1464	2	AAX30475
C 56	16.8	80.0	1554	2	AAT68140
C 57	16.8	80.0	2547	12	ADE91329
C 58	16.8	80.0	2953	6	ABV72307
C 59	16.8	80.0	48179	14	ADZ12700
C 60	16.8	80.0	51261	9	ADA02873

ALIGNMENTS

RESULT 1

ADD68803

ID ADD68803 standard; DNA; 21 BP.

XX AC ADD68803;

XX AC ADD68803;

DT 15-JAN-2004 (first entry)

XX PCR primer ptal used to amplify Group A Streptococcus ptal sequences.

DE beta-haemolytic Group A Streptococcus; GAS;

KW fluorescence resonance energy transfer; FRET; ptal; PCR; primer; ptal;

KW ss.

XX Streptococcus sp. 'group A'.

XX US6593093-B1.

PN 15-JUL-2003.

PD 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

PI

XX WPI; 2003-828265/77.  
XX Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.  
XX  
XX Claim 2; SEQ ID NO 2; 35pp; English.  
XX  
XX The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridising step comprising contacting the  
CC amplified product with fluorescent-labelled probes and finally, detecting  
CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the ptsI PCR primer of the invention which was used  
CC to amplify GAS ptsI sequences.  
XX  
XX Sequence 21 BP; 3 A; 4 C; 5 G; 9 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 21; DB 10; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
Db 1 TGCATGTATGGGTATCTTCC 21  
|||||

RESULT 2  
ADJ61990  
ID ADJ61990 standard; DNA; 21 BP.  
XX  
XX ADJ61990;  
XX  
XX 06-MAY-2004 (first entry)  
XX  
XX ptsI gene PCR primer, ptsI.  
XX  
XX Group A Streptococcus; GAS; PCR; primer; ss.  
XX  
XX Streptococcus sp.  
XX  
XX US2004014118-A1.  
XX  
XX 22-JAN-2004.  
XX  
XX 19-JUN-2003; 2003US-00465205.  
XX  
XX 20-FEB-2002; 2002US-00081923.  
XX  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
XX Uhl JR, Cockerill FR;  
XX  
XX WPI; 2004-224229/21.  
XX  
XX Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptsI primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX  
XX Claim 2; SEQ ID NO 2; 36pp; English.  
XX  
XX The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptsI primers and probes to produce ptsI  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of

CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a ptsI gene PCR primer.  
XX  
XX Sequence 21 BP; 3 A; 4 C; 5 G; 9 T; 0 U; 0 Other;  
SQ

Query Match 100.0%; Score 21; DB 12; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
Db 1 TGCATGTATGGGTATCTTCC 21  
|||||

RESULT 3  
ADZ75862  
ID ADZ75862 standard; DNA; 21 BP.  
XX  
XX ADZ75862;  
XX  
XX 28-JUL-2005 (first entry)  
XX  
XX Group B streptococcus phosphotransferase (pts) PCR primer, SEQ ID NO:2.  
XX  
XX Microorganism detection; fluorescence; diagnosis;  
KW streptococcus infection; infection; gynecology and obstetrics;  
KW phosphotransferase; PCR; primer; ss.  
XX  
XX Streptococcus sp. 'group B'.  
XX  
XX US2005106578-A1.  
XX  
XX 19-MAY-2005.  
XX  
XX 18-NOV-2003; 2003US-00716005.  
XX  
XX 18-NOV-2003; 2003US-00716005.  
XX  
XX {UHLJ/} UHL J R.  
XX {COCK/} COCKERILL F R.  
XX {AICH/} AICHINGER C.  
XX {REIS/} REISER A.  
XX  
XX Uhl JR, Cockerill FR, Aichinger C, Reiser A;  
XX  
XX WPI; 2005-371550/38.  
XX  
XX Detecting group B streptococcus, comprises amplifying a sample with pts  
PT primers, hybridizing the sample with fluorescently labeled pts probes,  
PT and detecting the presence of fluorescence resonance energy transfer.  
XX  
XX Claim 2; SEQ ID NO 2; 13pp; English.  
XX  
XX The invention relates to a real-time PCR-based method of detecting the  
CC presence or absence of group B streptococcus (GBS) bacterial pathogens in  
CC a biological sample from an individual. The method comprises  
CC amplification of a conserved region of the phosphotransferase (pts) gene  
CC of GBS (especially using PCR primers ADZ75861-ADZ75862), detection of the  
CC amplification product with a pair of fluorescently labeled  
CC phosphotransferase probes (especially ADZ75863-ADZ75864), and detection  
CC of the presence or absence of fluorescence resonance energy transfer  
CC (FRET), where presence of FRET is indicative of the presence of GBS in  
CC the sample. GBS infection is a leading cause of neonatal morbidity and  
CC mortality, with infection occurring during childbirth. Currently, it is  
CC recommended that women are screened for GBS during week 35-37 of  
CC gestation by a culture-based method which may take up to 72 hours for a  
CC result. However, many women first present at healthcare facilities at the  
CC time of labor, and in addition, GBS infection can be transient, so that a  
CC woman free of GBS at the time of screening may not be free of GBS when  
CC she is due to give birth. The method of the invention provides a real-

CC time assay for the detection of group B streptococcus in a sample, is  
CC more sensitive and specific than prior art non-culture based methods, and  
CC can thus be implemented for routine diagnosis of the presence of group B  
CC streptococcus. Sequences ADZ75861-ADZ75862 represent GBS  
CC phosphotransferase PCR primers specifically claimed for use in the method  
CC of the invention.

SQ Sequence 21 BP; 3 A; 4 C; 5 G; 9 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 14; Length 21;  
Best Local Similarity 100.0%; Pred. No. 1.3;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21  
Db 1 TGCATGTATGGGTTATCTTCC 21

## RESULT 4

ADD68796/C

ID ADD68796 standard; DNA; 1323 BP.

XX AC

ADD68796;

XX AC

15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' ptai DNA - SEQ ID 12.

DE beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptai; ds.

XX Streptococcus sp. 'group A'.

OS US6593093-B1.

XX FN

XX 15-JUL-2003.

XX 20-FEB-2002; 2002US-00081923.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2003-828265/77.

XX Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.

XX Example 2; SEQ ID NO 12; 35pp; English.

XX The invention relates to a novel method for detecting the presence or  
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
XX isolated from an individual by performing more than one cycling step. The  
XX method involves an amplifying step comprising contacting the sample with  
XX GAS primers, a subsequent hybridizing step comprising contacting the  
XX amplified product with fluorescent-labelled probes and finally, detecting  
XX fluorescence resonance energy transfer (FRET) where the presence or  
XX absence of FRET indicates the presence or absence of GAS in the sample.  
XX The method of the invention may be useful for detecting the presence or  
XX absence of GAS in a biological sample from an individual. The current  
XX sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 12)  
XX of the invention.

XX Sequence 1323 BP; 395 A; 278 C; 297 G; 353 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 21; DB 10; Length 1323;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21  
Db 193 TGCATGTATGGGTTATCTTCC 173

## RESULT 5

ADD62000/C

ID ADJ62000 standard; DNA; 1323 BP.

XX AC

ADJ62000;

XX DT

06-MAY-2004 (first entry)

XX Group A streptococcus ptai gene #7.

XX Group A Streptococcus; GAS; gene; ds.

XX Streptococcus sp.

XX US2004014118-A1.

XX 22-JAN-2004.

XX 19-JUN-2003; 2003US-00465205.

XX 20-FEB-2002; 2002US-00081923.

XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Uhl JR, Cockerill FR;

XX WPI; 2004-224229/21.

XX Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptai primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.

XX Disclosure; SEQ ID NO 12; 36pp; English.

XX The invention relates to a method of detecting the presence or absence of  
XX Group A Streptococcus (GAS) in biological sample from individual, by  
XX contacting sample with pair of ptai primers and probes to produce ptai  
XX amplification product, where first, second probes are labelled with donor  
XX and acceptor fluorescent moiety, and detecting presence or absence of  
XX fluorescence resonance energy transfer, which indicates the presence or  
XX absence of GAS in biological sample. The method is useful for detecting  
XX the presence or absence of Group A Streptococcus in a biological sample  
XX such as throat swabs, tissues and bodily fluids from an individual. The  
XX present sequence represents a group A streptococcus ptai gene.

XX Sequence 1323 BP; 395 A; 278 C; 297 G; 353 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 21; DB 12; Length 1323;

Best Local Similarity 100.0%; Pred. No. 2.1;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21  
Db 193 TGCATGTATGGGTTATCTTCC 173

## RESULT 6

ADD68798/C

ID ADD68798 standard; DNA; 1378 BP.

XX AC

ADD68798;

XX DT

15-JAN-2004 (first entry)

XX Streptococcus sp. 'group A' ptai DNA - SEQ ID 14.

DE beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptai; ds.

```
XX OS Streptococcus sp. 'group A'.
XX PN US6593093-B1.
XX PD 15-JUL-2003.
XX PF 20-FEB-2002; 2002US-00081923.
XX PR 20-FEB-2002; 2002US-00081923.
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX PI Uhl JR, Cockerill FR;
XX DR WPI; 2003-828265/77.
XX PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX PS Example 2; SEQ ID NO 14; 35pp; English.
XX CC The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridising step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptst DNA (SEQ ID 14)
XX of the invention.
XX SQ Sequence 1378 BP; 414 A; 289 C; 308 G; 367 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1378;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 7
ADJ62002/c
ID ADJ62002 standard; DNA; 1378 BP.
XX AC ADJ62002;
XX DT 06-MAY-2004 (first entry)
XX DE Group A streptococcus PstI gene #9.
XX KW Group A Streptococcus; GAS; gene; ds.
XX OS Streptococcus sp.
XX PN US2004014119-A1.
XX PD 22-JAN-2004.
XX PF 19-JUN-2003; 2003US-00465205.
XX PR 20-FEB-2002; 2002US-00081923.
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX PI Uhl JR, Cockerill FR;

Query Match 100.0%; Score 21; DB 12; Length 1378;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 8
ADD68800/c
ID ADD68800 standard; DNA; 1379 BP.
XX AC ADD68800;
XX DT 15-JAN-2004 (first entry)
XX DE Streptococcus sp. 'group A' ptst DNA - SEQ ID 16.
XX KW beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptst; ds.
XX OS Streptococcus sp. 'group A'.
XX PN US6593093-B1.
XX PD 15-JUL-2003.
XX PF 20-FEB-2002; 2002US-00081923.
XX PR 20-FEB-2002; 2002US-00081923.
XX PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX PI Uhl JR, Cockerill FR;
XX DR WPI; 2003-828265/77.
XX PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX PS Example 2; SEQ ID NO 16; 35pp; English.
XX CC The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridising step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptst DNA (SEQ ID 14)
XX of the invention.
XX SQ Sequence 1378 BP; 414 A; 289 C; 308 G; 367 T; 0 U; 0 Other;
```

CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 16)  
CC of the invention.  
XX  
SQ Sequence 1379 BP; 420 A; 291 C; 307 G; 361 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 21; DB 10; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 210 TGCATGTATGGTTATCTTCC 190  
  
RESULT 9  
ADD68797/c  
ID ADD68797 standard; DNA; 1379 BP.  
XX AC  
XX ADD68797;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Streptococcus sp. 'group A' ptaI DNA - SEQ ID 13.  
XX  
KW beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptaI; ds.  
XX  
OS Streptococcus sp. 'group A'.  
XX  
PN US6593093-B1.  
XX  
PD 15-JUL-2003.  
XX  
PF 20-FEB-2002; 2002US-00081923.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;  
XX  
DR WPI; 2003-828265/77.  
XX  
PT Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.  
XX  
PS Example 2; SEQ ID NO 13; 35pp; English.  
XX  
CC The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridising step comprising contacting the  
CC amplified product with fluorescent-labelled probes and finally, detecting  
CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 13)  
CC of the invention.  
XX  
SQ Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 21; DB 10; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 209 TGCATGTATGGTTATCTTCC 189  
  
RESULT 10  
ADD68791/c  
ID ADD68791 standard; DNA; 1379 BP.  
XX AC  
XX ADD68791;  
XX  
DT 15-JAN-2004 (first entry)  
XX  
DE Streptococcus sp. 'group A' ptaI DNA - SEQ ID 7.  
XX  
KW beta-haemolytic Group A Streptococcus; GAS;  
KW fluorescence resonance energy transfer; FRET; ptaI; ds.  
XX  
OS Streptococcus sp. 'group A'.  
XX  
PN US6593093-B1.  
XX  
PD 15-JUL-2003.  
XX  
PF 20-FEB-2002; 2002US-00081923.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;  
XX  
DR WPI; 2003-828265/77.  
XX  
PT Detecting presence/absence of group A Streptococcus in sample by  
PT performing cycling step comprising amplifying, hybridizing DNA from  
PT sample by fluorescent probe, detecting for fluorescent resonance energy  
PT transfer.  
XX  
PS Example 2; SEQ ID NO 7; 35pp; English.  
XX  
CC The invention relates to a novel method for detecting the presence or  
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
CC isolated from an individual by performing more than one cycling step. The  
CC method involves an amplifying step comprising contacting the sample with  
CC GAS primers, a subsequent hybridising step comprising contacting the  
CC amplified product with fluorescent-labelled probes and finally, detecting  
CC fluorescence resonance energy transfer (FRET) where the presence or  
CC absence of FRET indicates the presence or absence of GAS in the sample.  
CC The method of the invention may be useful for detecting the presence or  
CC absence of GAS in a biological sample from an individual. The current  
CC sequence is that of the Streptococcus sp. 'group A' ptaI DNA (SEQ ID 7)  
CC of the invention.  
XX  
SQ Sequence 1379 BP; 415 A; 289 C; 309 G; 366 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 21; DB 10; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TGCATGTATGGTTATCTTCC 21  
Db 209 TGCATGTATGGTTATCTTCC 189  
  
RESULT 11  
ADJ62001/c  
ID ADJ62001 standard; DNA; 1379 BP.  
XX AC  
XX ADJ62001;  
XX  
DT 06-MAY-2004 (first entry)  
XX

DE Group A streptococcus PtsI gene #8.  
XX  
KW Group A Streptococcus; GAS; gene; ds.  
XX  
OS Streptococcus sp.  
XX  
PN US2004014118-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 19-JUN-2003; 2003US-00465205.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;  
XX  
DR WPI; 2004-224229/21.  
XX  
PT Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptsI primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX  
PS Disclosure; SEQ ID NO 13; 36pp; English.  
XX  
CC The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptsI primers and probes to produce ptsI  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of  
CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a group A streptococcus ptsI gene.  
XX  
SQ Sequence 1379 BP; 413 A; 288 C; 309 G; 369 T; 0 U; 0 Other;  
Query Match 100.0%; Score 21; DB 12; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
RESULT 12  
ADJ62004/C  
ID ADJ62004 standard; DNA; 1379 BP.  
XX  
AC ADJ62004;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Group A streptococcus PtsI gene #11.  
XX  
KW Group A Streptococcus; GAS; gene; ds.  
XX  
OS Streptococcus sp.  
XX  
PN US2004014118-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 19-JUN-2003; 2003US-00465205.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;

DE Group A streptococcus PtsI gene #8.  
XX  
KW Group A Streptococcus; GAS; gene; ds.  
XX  
OS Streptococcus sp.  
XX  
PN US2004014118-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 19-JUN-2003; 2003US-00465205.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;

DE Group A streptococcus PtsI gene #8.  
XX  
KW Group A Streptococcus; GAS; gene; ds.  
XX  
OS Streptococcus sp.  
XX  
PN US2004014118-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 19-JUN-2003; 2003US-00465205.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;

XX  
DR WPI; 2004-224229/21.  
XX  
PT Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptsI primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX  
PS Disclosure; SEQ ID NO 16; 36pp; English.  
XX  
CC The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptsI primers and probes to produce ptsI  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of  
CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC such as throat swabs, tissues and bodily fluids from an individual. The  
CC present sequence represents a group A streptococcus ptsI gene.  
XX  
SQ Sequence 1379 BP; 420 A; 291 C; 307 G; 361 T; 0 U; 0 Other;  
Query Match 100.0%; Score 21; DB 12; Length 1379;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
RESULT 13  
ADJ61995/C  
ID ADJ61995 standard; DNA; 1379 BP.  
XX  
AC ADJ61995;  
XX  
DT 06-MAY-2004 (first entry)  
XX  
DE Group A streptococcus PtsI gene #2.  
XX  
KW Group A Streptococcus; GAS; gene; ds.  
XX  
OS Streptococcus sp.  
XX  
PN US2004014118-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 19-JUN-2003; 2003US-00465205.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;  
XX  
DR WPI; 2004-224229/21.  
XX  
PT Detecting Group A Streptococcus in biological sample, involves contacting  
PT sample with pair of ptsI primers and probes and detecting fluorescence  
PT resonance energy transfer indicating Group A Streptococcus in sample.  
XX  
PS Disclosure; SEQ ID NO 7; 36pp; English.  
XX  
CC The invention relates to a method of detecting the presence or absence of  
CC Group A Streptococcus (GAS) in biological sample from individual, by  
CC contacting sample with pair of ptsI primers and probes to produce ptsI  
CC amplification product, where first, second probes are labelled with donor  
CC and acceptor fluorescent moiety, and detecting presence or absence of  
CC fluorescence resonance energy transfer, which indicates the presence or  
CC absence of GAS in biological sample. The method is useful for detecting  
CC the presence or absence of Group A Streptococcus in a biological sample  
CC the presence or absence of Group A Streptococcus in a biological sample

DE Group A streptococcus PtsI gene #8.  
XX  
KW Group A Streptococcus; GAS; gene; ds.  
XX  
OS Streptococcus sp.  
XX  
PN US2004014118-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 19-JUN-2003; 2003US-00465205.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;

DE Group A streptococcus PtsI gene #8.  
XX  
KW Group A Streptococcus; GAS; gene; ds.  
XX  
OS Streptococcus sp.  
XX  
PN US2004014118-A1.  
XX  
PD 22-JAN-2004.  
XX  
PF 19-JUN-2003; 2003US-00465205.  
XX  
PR 20-FEB-2002; 2002US-00081923.  
XX  
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX  
PI Uhl JR, Cockerill FR;

CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus ptsI gene.

XX  
 SQ Sequence 1379 BP; 415 A; 289 C; 309 G; 366 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 21; DB 12; Length 1379;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
 |||||  
 Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 14  
 ADD68792/c  
 ID ADD68792 standard; DNA; 1384 BP.

XX  
 AC ADD68792;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Streptococcus sp. 'group A' ptsI DNA - SEQ ID 8.  
 XX  
 KW beta-haemolytic Group A Streptococcus; GAS;  
 KW fluorescence resonance energy transfer; FRET; ptsI; ds.  
 XX  
 OS Streptococcus sp. 'group A'.

XX  
 PN US6593093-B1.  
 XX  
 PD 15-JUL-2003.

XX  
 PF 20-FEB-2002; 2002US-00081923.  
 XX  
 PR 20-FEB-2002; 2002US-00081923.

XX  
 PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 PI Uhl JR, Cockerill FR;

XX  
 DR WPI; 2003-828265/77.  
 XX

XX  
 PT Detecting presence/absence of group A Streptococcus in sample by  
 PT performing cycling step comprising amplifying, hybridizing DNA from  
 PT sample by fluorescent probe, detecting for fluorescent resonance energy  
 PT transfer.

XX  
 PS Example 2; SEQ ID NO 8; 35pp; English.

XX  
 CC The invention relates to a novel method for detecting the presence or  
 CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
 CC isolated from an individual by performing more than one cycling step. The  
 CC method involves an amplifying step comprising contacting the sample with  
 CC GAS primers, a subsequent hybridising step comprising contacting the  
 CC amplified product with fluorescent-labelled probes and finally, detecting  
 CC fluorescence resonance energy transfer (FRET) where the presence or  
 CC absence of FRET indicates the presence or absence of GAS in the sample.  
 CC The method of the invention may be useful for detecting the presence or  
 CC absence of GAS in a biological sample from an individual. The current  
 CC sequence is that of the Streptococcus sp. 'group A' ptsI DNA (SEQ ID 8)  
 CC of the invention.

XX  
 SQ Sequence 1384 BP; 421 A; 291 C; 306 G; 366 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
 |||||  
 Db 210 TGCATGTATGGGTATCTTCC 190

RESULT 15  
 ADD68794/c

ID ADD68794 standard; DNA; 1384 BP.

XX  
 AC ADD68794;

XX  
 DT 15-JAN-2004 (first entry)

XX  
 DE Streptococcus sp. 'group A' ptsI DNA - SEQ ID 10.

XX  
 KW beta-haemolytic Group A Streptococcus; GAS;  
 KW fluorescence resonance energy transfer; FRET; ptsI; ds.

XX  
 OS Streptococcus sp. 'group A'.

XX  
 PN US6593093-B1.

XX  
 PD 15-JUL-2003.

XX  
 PF 20-FEB-2002; 2002US-00081923.

XX  
 PR 20-FEB-2002; 2002US-00081923.

XX  
 PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX  
 PI Uhl JR, Cockerill FR;

XX  
 DR WPI; 2003-828265/77.

XX  
 PT Detecting presence/absence of group A Streptococcus in sample by  
 PT performing cycling step comprising amplifying, hybridizing DNA from  
 PT sample by fluorescent probe, detecting for fluorescent resonance energy  
 PT transfer.

XX  
 PS Example 2; SEQ ID NO 10; 35pp; English.

XX  
 CC The invention relates to a novel method for detecting the presence or  
 CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
 CC isolated from an individual by performing more than one cycling step. The  
 CC method involves an amplifying step comprising contacting the sample with  
 CC GAS primers, a subsequent hybridising step comprising contacting the  
 CC amplified product with fluorescent-labelled probes and finally, detecting  
 CC fluorescence resonance energy transfer (FRET) where the presence or  
 CC absence of FRET indicates the presence or absence of GAS in the sample.  
 CC The method of the invention may be useful for detecting the presence or  
 CC absence of GAS in a biological sample from an individual. The current  
 CC sequence is that of the Streptococcus sp. 'group A' ptsI DNA (SEQ ID 10)  
 CC of the invention.

XX  
 SQ Sequence 1384 BP; 416 A; 292 C; 308 G; 368 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
 |||||  
 Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 16

ADJ61998/c

ID ADJ61998 standard; DNA; 1384 BP.

XX  
 AC ADJ61998;

XX  
 DT 06-MAY-2004 (first entry)

XX  
 DE Group A streptococcus PtsI gene #5.

XX  
 KW Group A Streptococcus; GAS; gene; ds.

OS Streptococcus sp.  
 XX US2004014118-A1.  
 XX 22-JAN-2004.  
 PD  
 XX  
 XX 19-JUN-2003; 2003US-00465205.  
 XX  
 XX 20-FEB-2002; 2002US-00081923.  
 PR  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 XX Uhl JR, Cockerill FR;  
 XX WPI; 2004-224229/21.  
 XX  
 XX Detecting Group A Streptococcus in biological sample, involves contacting  
 PT sample with pair of *ptsl* primers and probes and detecting fluorescence  
 PT resonance energy transfer indicating Group A Streptococcus in sample.  
 PT  
 XX Disclosure; SEQ ID NO 10; 36pp; English.  
 PS  
 XX The invention relates to a method of detecting the presence or absence of  
 CC Group A Streptococcus (GAS) in biological sample from individual, by  
 CC contacting sample with pair of *ptsl* primers and probes to produce *ptsl*  
 CC amplification product, where first, second probes are labelled with donor  
 CC and acceptor fluorescent moiety, and detecting presence or absence of  
 CC fluorescence resonance energy transfer, which indicates the presence or  
 CC absence of GAS in biological sample. The method is useful for detecting  
 CC the presence or absence of Group A Streptococcus in a biological sample  
 CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus *ptsl* gene.  
 CC  
 XX Sequence 1384 BP; 416 A; 292 C; 308 G; 368 T; 0 U; 0 Other;  
 SQ  
 Query Match 100.0%; Score 21; DB 12; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGCATGTATGGTTATCTTCC 21  
 Db 209 TGCATGTATGGTTATCTTCC 189  
 RESULT 17  
 ADJ61996/c  
 ID ADJ61996 standard; DNA; 1384 BP.  
 XX  
 AC ADJ61996;  
 XX  
 XX 06-MAY-2004 (first entry)  
 DT  
 XX Group A streptococcus *ptsl* gene #3.  
 DE  
 XX Group A Streptococcus; GAS; gene; ds.  
 KW  
 XX Streptococcus sp.  
 OS  
 XX US2004014118-A1.  
 XX  
 XX 22-JAN-2004.  
 PD  
 XX  
 XX 19-JUN-2003; 2003US-00465205.  
 XX  
 XX 20-FEB-2002; 2002US-00081923.  
 PR  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 XX Uhl JR, Cockerill FR;  
 XX WPI; 2004-224229/21.  
 XX  
 XX Detecting Group A Streptococcus in biological sample, involves contacting

PT sample with pair of *ptsl* primers and probes and detecting fluorescence  
 PT resonance energy transfer indicating Group A Streptococcus in sample.  
 XX Disclosure; SEQ ID NO 8; 36pp; English.  
 XX  
 XX The invention relates to a method of detecting the presence or absence of  
 CC Group A Streptococcus (GAS) in biological sample from individual, by  
 CC contacting sample with pair of *ptsl* primers and probes to produce *ptsl*  
 CC amplification product, where first, second probes are labelled with donor  
 CC and acceptor fluorescent moiety, and detecting presence or absence of  
 CC fluorescence resonance energy transfer, which indicates the presence or  
 CC absence of GAS in biological sample. The method is useful for detecting  
 CC the presence or absence of Group A Streptococcus in a biological sample  
 CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus *ptsl* gene.  
 CC  
 XX Sequence 1384 BP; 421 A; 291 C; 306 G; 366 T; 0 U; 0 Other;  
 SQ  
 Query Match 100.0%; Score 21; DB 12; Length 1384;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGCATGTATGGTTATCTTCC 21  
 Db 210 TGCATGTATGGTTATCTTCC 190  
 RESULT 18  
 ADD68793/c  
 ID ADD68793 standard; DNA; 1385 BP.  
 XX  
 AC ADD68793;  
 XX  
 XX 15-JAN-2004 (first entry)  
 DT  
 XX Streptococcus sp. 'group A' *ptsl* DNA - SEQ ID 9.  
 DE  
 XX beta-haemolytic Group A Streptococcus; GAS;  
 KW fluorescence resonance energy transfer; FRET; *ptsl*; ds.  
 XX  
 OS Streptococcus sp. 'group A'.  
 XX US6593093-B1.  
 PN  
 XX 15-JUL-2003.  
 PD  
 XX 20-FEB-2002; 2002US-00081923.  
 PF  
 XX 20-FEB-2002; 2002US-00081923.  
 PR  
 XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 XX Uhl JR, Cockerill FR;  
 PI  
 XX WPI; 2003-828265/77.  
 DR  
 XX  
 XX Detecting presence/absence of group A Streptococcus in sample by  
 PT performing cycling step comprising amplifying, hybridizing DNA from  
 PT sample by fluorescent probe, detecting for fluorescent resonance energy  
 PT transfer.  
 XX  
 XX Example 2; SEQ ID NO 9; 35pp; English.  
 PS  
 XX The invention relates to a novel method for detecting the presence or  
 CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
 CC isolated from an individual by performing more than one cycling step. The  
 CC method involves an amplifying step comprising contacting the sample with  
 CC GAS primers, a subsequent hybridizing step comprising contacting the  
 CC amplified product with fluorescent-labelled probes and finally, detecting  
 CC fluorescence resonance energy transfer (FRET) where the presence or  
 CC absence of FRET indicates the presence or absence of GAS in the sample.  
 CC The method of the invention may be useful for detecting the presence or  
 CC absence of GAS in a biological sample from an individual. The current



```
CC sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 9)
CC of the invention.
XX
SQ Sequence 1385 BP; 416 A; 293 C; 309 G; 367 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1385;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 19
ADJ61997/c
ID ADJ61997 standard; DNA; 1385 BP.
XX
AC ADJ61997;
XX
DT 06-MAY-2004 (first entry)
XX
DE Group A streptococcus ptai gene #4.
XX
KW Group A Streptococcus; GAS; gene; ds.
XX
OS Streptococcus sp.
XX
PN US2004014118-A1.
XX
PD 22-JAN-2004.
XX
PF 19-JUN-2003; 2003US-00465205.
XX
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
DR WPI; 2004-224229/21.
XX
PT Detecting Group A Streptococcus in biological sample, involves contacting
PT sample with pair of ptai primers and probes and detecting fluorescence
PT resonance energy transfer indicating Group A Streptococcus in sample.
XX
PS Disclosure; SEQ ID NO 9; 36pp; English.
XX
CC The invention relates to a method of detecting the presence or absence of
CC Group A Streptococcus (GAS) in biological sample from individual, by
CC contacting sample with pair of ptai primers and probes to produce ptai
CC amplification product, where first, second probes are labelled with donor
CC and acceptor fluorescent moiety, and detecting presence or absence of
CC fluorescence resonance energy transfer, which indicates the presence or
CC absence of GAS in biological sample. The method is useful for detecting
CC the presence or absence of Group A Streptococcus in a biological sample
CC such as throat swabs, tissues and bodily fluids from an individual. The
CC present sequence represents a group A streptococcus ptai gene.
XX
SQ Sequence 1385 BP; 416 A; 293 C; 309 G; 367 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 1385;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 210 TGCATGTATGGGTTATCTTCC 190

RESULT 20
ADD68795/c
ID ADD68795 standard; DNA; 1390 BP.
```

```
XX
AC ADD68795;
XX
DT 15-JAN-2004 (first entry)
XX
DE Streptococcus sp. 'group A' ptai DNA - SEQ ID 11.
XX
KW beta-haemolytic Group A Streptococcus; GAS;
KW fluorescence resonance energy transfer; FRET; ptai; ds.
XX
OS Streptococcus sp. 'group A'.
XX
PN US6593093-B1.
XX
PD 15-JUL-2003.
XX
PF 20-FEB-2002; 2002US-00081923.
XX
PR 20-FEB-2002; 2002US-00081923.
XX
PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
PI Uhl JR, Cockerill FR;
XX
DR WPI; 2003-828265/77.
XX
PT Detecting presence/absence of group A Streptococcus in sample by
PT performing cycling step comprising amplifying, hybridizing DNA from
PT sample by fluorescent probe, detecting for fluorescent resonance energy
PT transfer.
XX
PS Example 2; SEQ ID NO 11; 35pp; English.
XX
CC The invention relates to a novel method for detecting the presence or
CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
CC isolated from an individual by performing more than one cycling step. The
CC method involves an amplifying step comprising contacting the sample with
CC GAS primers, a subsequent hybridizing step comprising contacting the
CC amplified product with fluorescent-labelled probes and finally, detecting
CC fluorescence resonance energy transfer (FRET) where the presence or
CC absence of FRET indicates the presence or absence of GAS in the sample.
CC The method of the invention may be useful for detecting the presence or
CC absence of GAS in a biological sample from an individual. The current
CC sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 11)
CC of the invention.
XX
SQ Sequence 1390 BP; 419 A; 293 C; 310 G; 368 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1390;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
Db 224 TGCATGTATGGGTTATCTTCC 204

RESULT 21
ADJ61999/c
ID ADJ61999 standard; DNA; 1390 BP.
XX
AC ADJ61999;
XX
DT 06-MAY-2004 (first entry)
XX
DE Group A streptococcus ptai gene #6.
XX
KW Group A Streptococcus; GAS; gene; ds.
XX
OS Streptococcus sp.
XX
PN US2004014118-A1.
XX
```

```

PD 22-JAN-2004.
XX
XX
PF 19-JUN-2003; 2003US-00465205.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptsI primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 11; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptsI primers and probes to produce ptsI
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptsI gene.
XX
XX Sequence 1390 BP; 419 A; 293 C; 310 G; 368 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 21; DB 12; Length 1390;
XX Best Local Similarity 100.0%; Pred. No. 2.1;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TGCATGTATGGTTCCTCC 21
DB 224 TGCATGTATGGTTCCTCC 204

RESULT 22
ADD68799/c
ID ADD68799 standard; DNA; 1393 BP.
XX
XX AC ADD68799;
XX
XX 15-JAN-2004 (first entry)
XX
XX Streptococcus sp. 'group A' ptsI DNA - SEQ ID 15.
XX
XX beta-haemolytic Group A Streptococcus; GAS;
XX fluorescence resonance energy transfer; FRET; ptsI; ds.
XX
XX Streptococcus sp. 'group A'.
XX
XX US6593093-B1.
XX
XX 15-JUL-2003.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2003-828265/77.
XX
XX Detecting presence/absence of group A Streptococcus in sample by
XX performing cycling step comprising amplifying, hybridizing DNA from
XX sample by fluorescent probe, detecting for fluorescent resonance energy
XX transfer.
XX
XX Example 2; SEQ ID NO 15; 35pp; English.
XX
XX The invention relates to a novel method for detecting the presence or
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample
XX isolated from an individual by performing more than one cycling step. The
XX method involves an amplifying step comprising contacting the sample with
XX GAS primers, a subsequent hybridising step comprising contacting the
XX amplified product with fluorescent-labelled probes and finally, detecting
XX fluorescence resonance energy transfer (FRET) where the presence or
XX absence of FRET indicates the presence or absence of GAS in the sample.
XX The method of the invention may be useful for detecting the presence or
XX absence of GAS in a biological sample from an individual. The current
XX sequence is that of the Streptococcus sp. 'group A' ptsI DNA (SEQ ID 15)
XX of the invention.
XX
XX Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 21; DB 10; Length 1393;
XX Best Local Similarity 100.0%; Pred. No. 2.1;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TGCATGTATGGTTCCTCC 21
DB 210 TGCATGTATGGTTCCTCC 190

RESULT 23
ADD62003/c
ID ADJ62003 standard; DNA; 1393 BP.
XX
XX AC ADJ62003;
XX
XX 06-MAY-2004 (first entry)
XX
XX Group A streptococcus ptsI gene #10.
XX
XX Group A Streptococcus; GAS; gene; ds.
XX
XX Streptococcus sp.
XX
XX US2004014118-A1.
XX
XX 22-JAN-2004.
XX
XX 19-JUN-2003; 2003US-00465205.
XX
XX 20-FEB-2002; 2002US-00081923.
XX
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.
XX
XX Uhl JR, Cockerill FR;
XX
XX WPI; 2004-224229/21.
XX
XX Detecting Group A Streptococcus in biological sample, involves contacting
XX sample with pair of ptsI primers and probes and detecting fluorescence
XX resonance energy transfer indicating Group A Streptococcus in sample.
XX
XX Disclosure; SEQ ID NO 15; 36pp; English.
XX
XX The invention relates to a method of detecting the presence or absence of
XX Group A Streptococcus (GAS) in biological sample from individual, by
XX contacting sample with pair of ptsI primers and probes to produce ptsI
XX amplification product, where first, second probes are labelled with donor
XX and acceptor fluorescent moiety, and detecting presence or absence of
XX fluorescence resonance energy transfer, which indicates the presence or
XX absence of GAS in biological sample. The method is useful for detecting
XX the presence or absence of Group A Streptococcus in a biological sample
XX such as throat swabs, tissues and bodily fluids from an individual. The
XX present sequence represents a group A streptococcus ptsI gene.
XX
XX Sequence 1393 BP; 416 A; 293 C; 309 G; 375 T; 0 U; 0 Other;

```

Query Match 100.0%; Score 21; DB 12; Length 1393;  
 Best Local Similarity 100.0%; Pred. No. 2.1; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 DB 210 TGCATGTATGGGTTATCTTCC 190

RESULT 24  
 ADD6801/c  
 ID ADD6801 standard; DNA; 1447 BP.  
 XX  
 AC ADD6801;  
 XX  
 DT 15-JAN-2004 (first entry)  
 XX  
 DE Streptococcus sp. 'group A' pti DNA - SEQ ID 17.  
 XX  
 KW beta-haemolytic Group A Streptococcus; GAS;  
 KW fluorescence resonance energy transfer; FRET; pti; ds.  
 XX  
 OS Streptococcus sp. 'group A'.  
 XX  
 PN US6593093-B1.  
 XX  
 PD 15-JUL-2003.  
 XX  
 PF 20-FEB-2002; 2002US-00081923.  
 XX  
 PR 20-FEB-2002; 2002US-00081923.  
 XX  
 PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 PI Uhl JR, Cockerill FR;  
 XX  
 DR WPI; 2003-828265/77.  
 XX  
 PT Detecting presence/absence of group A Streptococcus in sample by  
 PT performing cycling step comprising amplifying, hybridizing DNA from  
 PT sample by fluorescent probe, detecting for fluorescent resonance energy  
 PT transfer.

XX  
 PS Example 2; SEQ ID NO 17; 35pp; English.  
 XX  
 CC The invention relates to a novel method for detecting the presence or  
 CC absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
 CC isolated from an individual by performing more than one cycling step. The  
 CC method involves an amplifying step comprising contacting the sample with  
 CC GAS primers, a subsequent hybridizing step comprising contacting the  
 CC amplified product with fluorescent-labelled probes and finally, detecting  
 CC fluorescence resonance energy transfer (FRET) where the presence or  
 CC absence of FRET indicates the presence or absence of GAS in the sample.  
 CC The method of the invention may be useful for detecting the presence or  
 CC absence of GAS in a biological sample from an individual. The current  
 CC sequence is that of the Streptococcus sp. 'group A' pti DNA (SEQ ID 17)  
 CC of the invention.

XX  
 SQ Sequence 1447 BP; 443 A; 304 C; 318 G; 382 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 10; Length 1447;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 DB 210 TGCATGTATGGGTTATCTTCC 190

RESULT 25  
 ADJ62005/c  
 ID ADJ62005 standard; DNA; 1447 BP.

XX  
 AC ADJ62005;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Group A streptococcus Pti gene #12.  
 XX  
 KW Group A Streptococcus; GAS; gene; ds.  
 XX  
 OS Streptococcus sp.  
 XX  
 PN US2004014118-A1.  
 XX  
 PD 22-JAN-2004.  
 XX  
 PF 19-JUN-2003; 2003US-00465205.  
 XX  
 PR 20-FEB-2002; 2002US-00081923.  
 XX  
 PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
 XX  
 PI Uhl JR, Cockerill FR;  
 XX  
 DR WPI; 2004-224229/21.  
 XX  
 CC Detecting Group A Streptococcus in biological sample, involves contacting  
 CC sample with pair of pti primers and probes and detecting fluorescence  
 CC resonance energy transfer indicating Group A Streptococcus in sample.  
 XX  
 PS Disclosure; SEQ ID NO 17; 36pp; English.  
 XX  
 CC The invention relates to a method of detecting the presence or absence of  
 CC Group A Streptococcus (GAS) in biological sample from individual, by  
 CC contacting sample with pair of pti primers and probes to produce pti  
 CC amplification product, where first, second probes are labelled with donor  
 CC and acceptor fluorescent moiety, and detecting presence or absence of  
 CC fluorescence resonance energy transfer, which indicates the presence or  
 CC absence of GAS in biological sample. The method is useful for detecting  
 CC the presence or absence of Group A Streptococcus in a biological sample  
 CC such as throat swabs, tissues and bodily fluids from an individual. The  
 CC present sequence represents a group A streptococcus pti gene.  
 XX  
 SQ Sequence 1447 BP; 443 A; 304 C; 318 G; 382 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 12; Length 1447;  
 Best Local Similarity 100.0%; Pred. No. 2.2; Mismatches 0; Indels 0; Gaps 0;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
 DB 210 TGCATGTATGGGTTATCTTCC 190

RESULT 26  
 ABN68277/c  
 ID ABN68277 standard; DNA; 1731 BP.  
 XX  
 AC ABN68277;  
 XX  
 DT 01-JUL-2002 (first entry)  
 XX  
 DE Streptococcus polynucleotide SEQ ID NO 4467.  
 XX  
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX  
 OS Streptococcus pyogenes.  
 XX  
 PN WO200234771-A2.  
 XX  
 PD 02-MAY-2002.  
 XX

PF 29-OCT-2001; 2001WO-GB004789.  
XX  
PR 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
DR P-PSDB; ABP27646.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
PT  
XX  
XX Claim 7; Page 3608; 4525pp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus (I) may be prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
XX Sequence 1731 BP; 517 A; 358 C; 389 G; 467 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 21; DB 6; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
Db 377 TGCATGTATGGGTATCTTCC 357  
RESULT 27  
ABN68276/c  
ID ABN68276 standard; DNA; 1731 BP.  
XX  
AC ABN68276;  
XX  
XX 01-JUL-2002 (first entry)  
XX  
XX Streptococcus polynucleotide SEQ ID NO 4465.  
DE  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX  
XX Streptococcus agalactiae.  
OS  
XX WO200234771-A2.  
PN  
XX 02-MAY-2002.  
PD  
XX 29-OCT-2001; 2001WO-GB004789.  
PF  
XX 27-OCT-2000; 2000GB-00026333.  
PR

PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Masignani V, Margarit Y Rosi, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
DR P-PSDB; ABP27645.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
PT  
XX  
XX Claim 7; Page 3607; 4525pp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus (I) may be prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
XX Sequence 1731 BP; 541 A; 323 C; 373 G; 494 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 21; DB 6; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGGTATCTTCC 21  
Db 377 TGCATGTATGGGTATCTTCC 357  
RESULT 28  
ACA50572/c  
ID ACA50572 standard; DNA; 1734 BP.  
XX  
AC ACA50572;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Prokaryotic essential gene #32229.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX Streptococcus pyogenes.  
XX  
XX WO200277183-A2.  
PN  
XX 03-OCT-2002.  
PD  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
PR

XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX PI WPI; 2003-029926/02.  
XX DR P-PSDB; ABU46702.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX PS Claim 14; SEQ ID NO 38442; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
XX prokaryotic essential genes. Note: The sequence data for this patent did  
XX not form part of the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX SQ Sequence 1734 BP; 519 A; 358 C; 389 G; 468 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 21; DB 8; Length 1734;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 377 TGCATGTATGGGTATCTTCC 357  
  
RESULT 29  
ADV85417/c  
ID ADV85417 standard; DNA; 1734 BP.  
XX AC ADV85417;  
XX  
XX DT 24-FEB-2005 (first entry)  
XX  
XX DE Streptococcus agalactiae DNA sequence, SEQ ID 6558.  
XX  
XX KW Antibacterial; vaccine; bacterial infection; ds.  
XX  
XX OS Streptococcus agalactiae.  
XX  
XX PS WO200292819-A2.

XX PD 21-NOV-2002.  
XX PF 26-APR-2002; 2002WO-1B003059.  
XX PR 26-APR-2001; 2001PR-00005642.  
XX PA (INSP ) INST PASTEUR.  
XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX PI Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalloui L;  
XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
XX WPI; 2004-101891/11.  
XX  
XX Genomic nucleotide sequences encoding polypeptides of *Streptococcus*  
XX *agalactiae* for the development of vaccines, diagnostic tools, DNA chips  
XX and identification of therapeutic targets.  
XX  
XX PS Claim 4; SEQ ID NO 6558; 439pp; French.  
XX  
XX The present invention relates to novel *Streptococcus agalactiae*  
XX nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV83476) and  
XX novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
XX nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
XX synthesis of amino acids, cell membranes, intermediate (central)  
XX metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
XX nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
XX regulatory functions, replication, transcription, translation, protein  
XX transport, adaptation to atypical conditions, sensitivity to medicines  
XX and/or analogues, functions related to transposons, biosynthesis of  
XX cofactors, prosthetic groups and transporters, cell membrane proteins and  
XX cellular machinery. (I) are useful for the detection and/or amplification  
XX of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
XX useful for treatment of a bacterial *S. agalactiae* infection. The complete  
XX genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
XX present patent is an equivalent for the basic patent FR2824074A1, which  
XX contains only 2344 sequences.  
XX  
XX SQ Sequence 1734 BP; 540 A; 323 C; 376 G; 495 T; 0 U; 0 Other;  
  
Query Match 100.0%; Score 21; DB 13; Length 1734;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 377 TGCATGTATGGGTATCTTCC 357  
  
RESULT 30  
ADD68806/c  
ID ADD68806 standard; DNA; 1803 BP.  
XX AC ADD68806;  
XX  
XX DT 15-JAN-2004 (first entry)  
XX  
XX DE *Streptococcus* sp. 'group A' ptaI DNA - SEQ ID 5.  
XX  
XX KW beta-haemolytic Group A *Streptococcus*; GAS;  
XX KW fluorescence resonance energy transfer; FRET; ptaI; ds; gene.  
XX  
XX OS *Streptococcus* sp. 'group A'.  
XX  
XX FH Key Location/Qualifiers  
XX FT CDS 1..1803  
XX FT /\*tag= a  
XX FT /partial  
XX FT /product= "Streptococcus sp. 'group A' ptaI protein - SEQ  
XX FT ID 6"  
XX FT /note= "No stop codon; this CDS contains translation  
XX FT exceptions"

XX US6593093-B1.  
XX 15-JUL-2003.  
XX 20-FEB-2002; 2002US-00081923.  
XX 20-FEB-2002; 2002US-00081923.  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX Uhl JR, Cockerill FR;  
XX WPI; 2003-828265/77.  
XX P-PSDB; ADD68790.  
XX  
XX Detecting presence/absence of group A Streptococcus in sample by  
XX performing cycling step comprising amplifying, hybridizing DNA from  
XX sample by fluorescent probe, detecting for fluorescent resonance energy  
XX transfer.  
XX  
XX Example 2; SEQ ID NO 5; 35pp; English.  
XX  
XX The invention relates to a novel method for detecting the presence or  
XX absence of beta-haemolytic Group A Streptococcus (GAS) in a sample  
XX isolated from an individual by performing more than one cycling step. The  
XX method involves an amplifying step comprising contacting the sample with  
XX GAS primers, a subsequent hybridizing step comprising contacting the  
XX amplified product with fluorescent-labelled probes and finally, detecting  
XX fluorescence resonance energy transfer (FRET) where the presence or  
XX absence of FRET indicates the presence or absence of GAS in the sample.  
XX The method of the invention may be useful for detecting the presence or  
XX absence of GAS in a biological sample from an individual. The current  
XX sequence is that of the Streptococcus sp. 'group A' ptai DNA (SEQ ID 5)  
XX  
XX  
XX Sequence 1803 BP; 540 A; 371 C; 406 G; 486 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 21; DB 10; Length 1803;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGTTATCTTCC 21  
Db 377 TGCATGTATGGTTATCTTCC 357  
RESULT 31  
ADJ61993/C  
ID ADJ61993 standard; DNA; 1803 BP.  
XX AC ADJ61993;  
XX 06-MAY-2004 (first entry)  
XX Group A streptococcus PtaI gene #1.  
XX Group A Streptococcus; GAS; gene; ds.  
XX Streptococcus sp.  
XX Key Location/Qualifiers  
XX CDS 1..1803  
XX /\*tag= a  
XX /product= "PtaI"  
XX /note= "This sequence contains translation exceptions"  
XX  
XX US2004014118-A1.  
XX 22-JAN-2004.  
XX 19-JUN-2003; 2003US-00465205.  
XX

PR 20-FEB-2002; 2002US-00081923.  
XX (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.  
XX Uhl JR, Cockerill FR;  
XX WPI; 2004-224229/21.  
XX P-PSDB; ADJ61994.  
XX  
XX Detecting Group A Streptococcus in biological sample, involves contacting  
XX sample with pair of ptai primers and probes and detecting fluorescence  
XX resonance energy transfer indicating Group A Streptococcus in sample.  
XX  
XX Disclosure; SEQ ID NO 5; 36pp; English.  
XX  
XX The invention relates to a method of detecting the presence or absence of  
XX Group A Streptococcus (GAS) in biological sample from individual, by  
XX contacting sample with pair of ptai primers and probes to produce ptai  
XX amplification product, where first, second probes are labelled with donor  
XX and acceptor fluorescent moiety, and detecting presence or absence of  
XX fluorescence resonance energy transfer, which indicates the presence or  
XX absence of GAS in biological sample. The method is useful for detecting  
XX the presence or absence of Group A Streptococcus in a biological sample  
XX such as throat swabs, tissues and bodily fluids from an individual. The  
XX present sequence represents a group A streptococcus ptai gene.  
XX  
XX Sequence 1803 BP; 540 A; 371 C; 406 G; 486 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 21; DB 12; Length 1803;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGTTATCTTCC 21  
Db 377 TGCATGTATGGTTATCTTCC 357  
RESULT 32  
ADV87723/C  
ID ADV87723 standard; DNA; 29072 BP.  
XX AC ADV87723;  
XX 24-FEB-2005 (first entry)  
XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
XX Antibacterial; Vaccine; bacterial infection; ds.  
XX Streptococcus agalactiae.  
XX FR2824074-A1.  
XX 31-OCT-2002.  
XX 26-APR-2001; 2001FR-00005642.  
XX 26-APR-2001; 2001FR-00005642.  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX Glaeser P, Rusniok C, Chevalier F, Frangeul L, Lalloui L;  
XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;  
XX WPI; 2004-101891/11.  
XX  
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX and identification of therapeutic targets.  
XX  
XX Claim 1; SEQ ID NO 117; 2687pp; French.  
XX

CC The present invention relates to novel Streptococcus agalactiae  
CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
CC agalactiae involved in the synthesis of amino acids, cell membranes,  
CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
CC phospholipid metabolism, nucleotide metabolism including purines,  
CC pyrimidines and/or nucleosides, regulatory functions, replication,  
CC transcription, translation, protein transport, adaptation to atypical  
CC conditions, sensitivity to medicines and/or analogues, functions related  
CC to transposons, biosynthesis of cofactors, prosthetic groups and  
CC transporters, cell membrane proteins and cellular machinery. (I) are  
CC useful for the detection and/or amplification of nucleic acids.  
CC Pharmaceutical composition comprising (I) or (II) are useful for  
CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
CC contains 6617 sequence whereas the present patent only contains 2344  
CC sequences.  
XX  
SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTTCC 21  
Db 12272 TGCATGATGGGTATCTTCC 12252

## RESULT 33

ADV78976/c

ID ADV78976 standard; DNA; 29072 BP.

XX AC ADV78976;

XX XX 24-FEB-2005 (first entry)

XX DE Streptococcus agalactiae DNA sequence, SEQ ID 117.

XX XX Antibacterial; vaccine; bacterial infection; ds.

XX XX Streptococcus agalactiae.

XX OS WO200292818-A2.

XX XX 21-NOV-2002.

XX PF 26-APR-2002; 2002WO-IB003059.

XX PR 26-APR-2001; 2001FR-00005642.

XX XX (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Glaser P, Rusniok C, Chevalier P, Frangeul L, Lalioui L;

XX PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;

XX XX WPI; 2004-101891/11.

XX XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX PT and identification of therapeutic targets.

XX PS Claim 1; SEQ ID NO 117; 439pp; French.

XX CC The present invention relates to novel Streptococcus agalactiae  
XX CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
XX CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
XX CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
XX CC synthesis of amino acids, cell membranes, intermediate (central)  
XX CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
XX CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
XX CC regulatory functions, replication, transcription, translation, protein

CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (I) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
CC useful for treatment of a bacterial S. agalactiae infection. The complete  
CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.

XX SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

Query Match 100.0%; Score 21; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTTCC 21  
Db 12272 TGCATGATGGGTATCTTCC 12252

## RESULT 34

ABN71527\_07/c

WP Continuation (8 of 22) of ABN71527 from base 700001 (Streptococcus polynucleotide SEQ 1  
WP Sequence split into 22 fragments LOCUS ABN71527 Accession ABN71527

WP	Fragment Name	Begin	End
WP	ABN71527_00	1	110000
WP	ABN71527_01	100001	210000
WP	ABN71527_02	200001	310000
WP	ABN71527_03	300001	410000
WP	ABN71527_04	400001	510000
WP	ABN71527_05	500001	610000
WP	ABN71527_06	600001	710000
WP	ABN71527_07	700001	810000
WP	ABN71527_08	800001	910000
WP	ABN71527_09	900001	1010000
WP	ABN71527_10	1000001	1110000
WP	ABN71527_11	1100001	1210000
WP	ABN71527_12	1200001	1310000
WP	ABN71527_13	1300001	1410000
WP	ABN71527_14	1400001	1510000
WP	ABN71527_15	1500001	1610000
WP	ABN71527_16	1600001	1710000
WP	ABN71527_17	1700001	1810000
WP	ABN71527_18	1800001	1910000
WP	ABN71527_19	1900001	2010000
WP	ABN71527_20	2000001	2110000
WP	ABN71527_21	2100001	2155561

Query Match 100.0%; Score 21; DB 6; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 3.6;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTTCC 21  
Db 91142 TGCATGATGGGTATCTTCC 91122

## RESULT 35

ADV81204\_08/c

WP Continuation (9 of 23) of ADV81204 from base 800001 (Streptococcus agalactiae complete  
WP Sequence split into 23 fragments LOCUS ADV81204 Accession ADV81204

WP	Fragment Name	Begin	End
WP	ADV81204_00	1	110000
WP	ADV81204_01	100001	210000
WP	ADV81204_02	200001	310000
WP	ADV81204_03	300001	410000
WP	ADV81204_04	400001	510000
WP	ADV81204_05	500001	610000
WP	ADV81204_06	600001	710000
WP	ADV81204_07	700001	810000
WP	ADV81204_08	800001	910000
WP	ADV81204_09	900001	1010000

```
WP ADV81204_10 1110000 1000001 1110000
WP ADV81204_11 1100001 1210000
WP ADV81204_12 1200001 1310000
WP ADV81204_13 1300001 1410000
WP ADV81204_14 1400001 1510000
WP ADV81204_15 1500001 1610000
WP ADV81204_16 1600001 1710000
WP ADV81204_17 1700001 1810000
WP ADV81204_18 1800001 1910000
WP ADV81204_19 1900001 2010000
WP ADV81204_20 2000001 2110000
WP ADV81204_21 2100001 2210000
WP ADV81204_22 2200001 2217924

Query Match 100.0%; Score 21; DB 13; Length 110000;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
Db 71477 TGCATGTATGGGTATCTTCC 71457

RESULT 36
ABX06888/c
ID ABX06888 standard; DNA; 402 BP.
AC ABX06888;
XX
XX 27-OCT-2003 (revised)
XX 11-FEB-2003 (first entry)
XX
XX S. pneumoniae type 4 strain coding region #1176.
XX
XX Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
XX ear infection; antiinflammatory; antibacterial; immunostimulant;
XX auditory; respiratory; gene therapy; vaccine.
XX
XX Streptococcus pneumoniae; type 4 strain.
XX
XX WO200277021-A2.
XX
XX 03-OCT-2002.
XX
XX 27-MAR-2002; 2002WO-IB002163.
XX
XX 27-MAR-2001; 2001GB-00007658.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Masignani V, Tettelin H, Fraser C;
XX
XX WPI; 2003-040579/03.
XX P-PSDB; ABU01600.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX useful as medicaments for treating or preventing a disease or infection
XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
XX ear infection.
XX
XX Claim 6; SEQ ID NO 2351; 56pp; English.
XX
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX AB556454. Also included are an antibody which binds one of the proteins,
XX treating a patient by administering the protein, DNA or antibody (in a
XX composition), a kit comprising first and second primers, which are the
XX nucleic acid cited above or fragments between nucleotides 8-100 of a
XX sequence not defined in the specification, for amplifying a target
```

```
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the target sequence
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2489
CC identified coding region from the genomic sequence. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences. (Updated on 27-OCT-2003 to
CC standardise OS field)
XX
XX Sequence 402 BP; 128 A; 85 C; 96 G; 93 T; 0 U; 0 Other;
XX
XX Query Match 92.4%; Score 19.4; DB 10; Length 402;
XX Best Local Similarity 95.2%; Pred. No. 12;
XX Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 TGCATGTATGGGTATCTTCC 21
XX |||||||||||
XX Db 377 TGCATGTATGGGTATCTTCC 357

RESULT 37
AAV65250/c
ID AAV65250 standard; DNA; 962 BP.
XX
XX AAV65250;
XX
XX 24-DEC-1998 (first entry)
XX
XX DNA encoding a S. pneumoniae PRP-protein PTS enzyme I.
XX
XX Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;
XX virulence; antibody; infection; detection; treatment; ss.
XX
XX Streptococcus pneumoniae.
XX
XX WO9826072-A1.
XX
XX 18-JUN-1998.
XX
XX 09-DEC-1997; 97WO-US022578.
XX
XX 13-DEC-1996; 96US-0036281P.
XX
XX (ELIL) LILLY & CO ELI.
XX
XX Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jaskunas SR;
XX Mills BJ, Norris FH, Peery RB, Rostek PK, Rostek PR, Skatrud PL;
XX Smith MC, Solenberg PJ, Treadway PJ, Young Bellido ML;
XX
XX WPI; 1998-348529/30.
XX P-PSDB; AAW80678.
XX
XX Streptococcus pneumoniae nucleic acid sequences - used in DNA chips for
XX evaluating gene expression, and identification of virulence genes.
XX
XX Claim 1; Page 105; 333pp; English.
XX
XX This DNA sequence encodes a Streptococcus pneumoniae PRP-protein PTS
XX enzyme I. The invention provides DNA sequences (AAV65201 to AAV65304)
XX from the Streptococcus pneumoniae genome and corresponding protein
XX sequences (AAW80605 to AAW80728). A recombinant host containing a vector
```



CC comprising any of the above nucleic acids can be used for the recombinant  
CC expression of the protein sequences. The invention also provides a DNA  
CC chip having arrayed on it at least 15 base pair fragment of any one or  
CC more of these DNA sequences. The DNA chip can be used methods for  
CC evaluating gene expression in *S. pneumoniae* and for identifying virulence  
CC genes in *S. pneumoniae*. Antibodies that selectively bind to the above  
CC proteins or peptide fragments can be used to treat *S. pneumoniae*  
CC infection. The antibodies can also be used to detect *S. pneumoniae* cells  
XX  
SQ Sequence 962 BP; 312 A; 196 C; 224 G; 230 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 2; Length 962;  
Best Local Similarity 95.2%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTCC 21  
Db 519 TGCATGTATGGTTATCTTCC 499

RESULT 38  
ADK44676/c  
ID ADK44676 standard; DNA; 1140 BP.

AC ADK44676;

XX 24-FEB-2005 (first entry)

DT Streptococcus pneumoniae gene, Seq ID No 1191.

DE db; gene; Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.

KW Streptococcus pneumoniae.

OS US6699703-B1.

PN 02-MAR-2004.

PD 26-MAY-2000; 2000US-00583110.

XX 02-JUL-1997; 97US-0051553P.

PR 12-MAY-1998; 98US-0085131P.

PR 30-JUN-1998; 98US-00107433.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;

XX WPI; 2004-212399/20.

DR P-PSDB; ADK47337.

XX New nucleic acid molecules and polypeptides useful for diagnosing,  
PT preventing and treating pathological conditions resulting from bacterial  
PT infection, e.g. Streptococcus pneumoniae infection, and in drug  
PT screening.

XX Disclosure; SEQ ID NO 1191; 301pp; English.

XX The invention relates to isolated Streptococcus pneumoniae nucleic acids  
CC and polypeptides. The nucleic acids and proteins are useful for  
CC diagnosing, preventing and treating pathological conditions resulting  
CC from bacterial infection, such as *S. pneumoniae* infection. These may also  
CC be used for drug screening procedures. The present sequence represents a  
CC Streptococcus pneumoniae nucleic acid of the invention. Note: The  
CC sequence data for this patent did not appear in the printed specification  
CC but was obtained in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html.

XX Sequence 1140 BP; 355 A; 239 C; 259 G; 287 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 13; Length 1140;  
Best Local Similarity 95.2%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTCC 21  
Db 377 TGCATGTATGGTTATCTTCC 357

RESULT 39  
ADR92834/c

ID ADR92834 standard; DNA; 1287 BP.

XX ADR92834;

XX 16-DEC-2004 (first entry)

XX Novel *S. pneumoniae* DNA sequence, SEQ ID 1469.

XX Meningitis; bacteraemia; pneumonia; otitis media; da;  
KW bacterial infection.

XX Streptococcus pneumoniae.

XX US6800744-B1.

XX 05-OCT-2004.

XX 30-JUN-1998; 98US-00107433.

XX 02-JUL-1997; 97US-0051553P.

PR 12-MAY-1998; 98US-0085131P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2004-697205/68.

DR P-PSDB; ADR95437.

XX New isolated nucleic acid encoding a Streptococcus pneumoniae  
PT polypeptide useful for diagnosing, preventing and/or treating  
PT pathological conditions resulting from the bacterial infection.

XX Disclosure; SEQ ID NO 1469; 151pp; English.

XX The invention relates to an isolated nucleic acid comprising a sequence  
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its  
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,  
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,  
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,  
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92850 or  
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide  
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is  
CC hybridisable under high stringency conditions to the nucleotide sequence.  
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.  
CC Also included are a recombinant expression vector comprising the isolated  
CC nucleic acid cited above operably linked to a transcription regulatory  
CC element, a cell comprising the recombinant expression vector and a probe  
CC comprising at least 20 consecutive nucleotides of the nucleotide  
CC sequences as cited above. The methods and compositions of the present  
CC invention are useful for the diagnosis, prevention and/or treatment of  
CC pathological conditions resulting from bacterial infection by  
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and  
CC otitis media. The present sequence is one of the 2603 disclosed *S.*  
CC pneumoniae nucleic acid sequences. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.

XX Sequence 1287 BP; 396 A; 271 C; 295 G; 325 T; 0 U; 0 Other;

Query Match 92.4%; Score 19.4; DB 13; Length 1287;  
Best Local Similarity 95.2%; Pred. No. 13;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
OY 1 TGCATGTATGGGTTATCTTCC 21
    |||||
Db 431 TGCATGTATGGGTTGCTTCC 411

RESULT 40
ID AEA56704/c
XX AEA56704 standard; DNA; 1287 BP.
AC AEA56704;
XX
XX
XX 25-AUG-2005 (first entry)
XX
XX Streptococcus pneumoniae ORF nucleic acid sequence SEQ ID NO:1469.
XX
XX bacterial infection; Streptococcus pneumoniae infection; antibacterial;
XX vaccine; gene; ds.
XX
XX Streptococcus pneumoniae.
XX
XX US2005136404-A1.
XX
XX 23-JUN-2005.
XX
XX 10-JUL-2003; 2003US-00617320.
XX
XX 02-JUL-1997; 97US-0051553P.
XX 12-MAY-1998; 98US-0085131P.
XX 30-JUN-1998; 98US-00107433.
XX
XX (DOUC/) DOUCETTE-STAMM L A.
XX (BUSH/) BUSH D.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2005-477576/48.
XX P-PSDB; AEA59307.
XX
XX New isolated nucleic acid molecules and encoded polypeptides useful for
XX diagnosing, preventing or treating bacterial infections, particularly
XX Streptococcus pneumoniae infection.
XX
XX Claim 1; SEQ ID NO 1469; 144pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule for detecting,
XX preventing or treating pathological conditions resulting from bacterial
XX infection. The isolated nucleic acid comprises: (a) any of the 2603
XX nucleotide sequences of AEA5236 to AEA57838; (b) a nucleotide sequence
XX encoding a Streptococcus pneumoniae polypeptide comprising any of the
XX 2603 amino acid sequences of AEA57839 to AEA6041; or (c) a nucleotide
XX sequence of at least 8 nucleotides in length, where the sequence is
XX hybridizable to a nucleic acid having any of the nucleotide sequences in
XX (a). Also described: (1) a recombinant expression vector comprising the
XX above nucleic acid operably linked to a transcription regulatory element;
XX (2) a cell comprising the recombinant expression vector; (3) producing an
XX S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
XX consisting of at least 8 nucleotides of any of AEA5236 to AEA57838; (5)
XX treating a subject for S. pneumoniae infection; (6) a recombinant or
XX substantially pure preparation of an S. pneumoniae polypeptide or its
XX fragment, where the polypeptide is selected from AEA57839 to AEA6041;
XX (7) a vaccine composition for preventing or treating an S. pneumoniae
XX infection, comprising an amount of the above nucleic acid or polypeptide;
XX (8) detecting the presence of a Streptococcus nucleic acid in a sample;
XX (9) a computer readable medium having recorded the nucleotide sequences
XX of AEA5236 to AEA57838; (10) a computer based system for identifying
XX fragments of the Streptococcus genome of commercial importance. The
XX composition and methods are useful for diagnosing, preventing or treating
XX bacterial infections, particularly S. pneumoniae infection. The present
XX sequence represents a S. pneumoniae ORF nucleic acid sequence from the
XX present invention. Note - The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from the USPTO web site.
XX

SQ Sequence 1287 BP; 396 A; 271 C; 295 G; 325 T; 0 U; 0 Other;
Query Match 92.4%; Score 19.4; DB 14; Length 1287;
Best Local Similarity 95.2%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCATGTATGGGTTATCTTCC 21
    |||||
Db 431 TGCATGTATGGGTTGCTTCC 411

RESULT 41
ID AAS55546/c
XX AAS55546 standard; DNA; 1734 BP.
AC AAS55546;
XX
XX 13-FEB-2002 (first entry)
XX
XX Streptococcus pneumoniae DNA for cellular proliferation protein #117.
XX
XX Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
XX antibacterial; drug design.
XX
XX Streptococcus pneumoniae.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US009180.
XX
XX 21-MAR-2000; 2000US-0191078P.
XX 23-MAY-2000; 2000US-0206848P.
XX 26-MAY-2000; 2000US-0207727P.
XX 23-OCT-2000; 2000US-0242578P.
XX 27-NOV-2000; 2000US-0253625P.
XX 22-DEC-2000; 2000US-0257931P.
XX 16-FEB-2001; 2001US-0269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX P-PSDB; AAU37687.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX
XX Claim 27; SEQ ID NO 9183; 51pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX their use in the discovery of novel antibiotics, the essential genes
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX nucleic acids which are required for cell proliferation in a wide variety
XX of organisms. The present sequence encodes an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1734 BP; 531 A; 365 C; 390 G; 448 T; 0 U; 0 Other;
```

Query Match 92.4%; Score 19.4; DB 4; Length 1734;  
 Best Local Similarity 95.2%; Pred. No. 14;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTCCTCC 21  
 |||||  
 Db 377 TGCATGTATGGTTCCTCC 357

RESULT 42  
 ACA49934/C  
 ID ACA49934 standard; DNA; 1734 BP.  
 XX  
 AC ACA49934;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #31591.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 XX drug design; gene.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO20027183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 XX WPI: 2003-029926/02.  
 DR P-PSDB; ABU46064.  
 DR  
 XX  
 PS New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 PS Claim 14; SEQ ID NO 37804; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for

CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1734 BP; 531 A; 365 C; 390 G; 448 T; 0 U; 0 Other;  
 SQ  
 Query Match 92.4%; Score 19.4; DB 8; Length 1734;  
 Best Local Similarity 95.2%; Pred. No. 14;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTCCTCC 21  
 |||||  
 Db 377 TGCATGTATGGTTCCTCC 357

RESULT 43  
 AAV52227/C  
 ID AAV52227 standard; DNA; 8195 BP.  
 XX  
 AC AAV52227;  
 XX  
 DT 23-OCT-1998 (first entry)  
 XX  
 DE Streptococcus pneumoniae genome fragment SEQ ID NO:94.  
 XX  
 KW Streptococcus pneumoniae; *S. pneumoniae*; genome; diagnosis; assay;  
 KW computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX  
 OS Streptococcus pneumoniae.  
 XX  
 PN WO9818931-A2.  
 XX  
 PD 07-MAY-1998.  
 XX  
 PF 30-OCT-1997; 97WO-US019588.  
 XX  
 PR 31-OCT-1996; 96US-0029960P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;  
 PI Dougherty BA;  
 XX  
 XX WPI: 1998-272225/24.  
 XX  
 XX Computer-readable medium with recorded Streptococcus pneumoniae  
 PT polynucleotide sequences - useful in diagnostic kits and assays, and  
 PT pharmaceutical compositions and vaccines for Streptococcus pneumoniae.  
 XX  
 PS Claim 1; Page 727-732; 1409pp; English.

XX The present invention describes a computer readable medium which has the  
 CC nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
 CC on it, or a representative fragment or a sequence at least 95% identical  
 CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 CC to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
 CC pneumoniae. The present invention also describes an isolated nucleic acid  
 CC molecule encoding a homologue of any of the fragments of the *S. pneumoniae*  
 CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 CC by a process comprising: (a) screening a genomic DNA library using as a  
 CC probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
 CC 391, identifying members of the library which contain sequences that  
 CC hybridise to the target sequence and isolating the nucleic acid molecules  
 CC from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
 CC organism, amplifying nucleic acid molecules whose nucleotide sequence is  
 CC homologous to amplification primers derived from the fragment of the *S.*  
 CC pneumoniae genome to prime the amplification and isolating the amplified

CC sequences. The computer readable medium can be used in a computer-based  
CC system for identifying fragments of the *S. pneumoniae* genome of  
CC commercial importance, or expression modulating fragments of the *S.*  
CC *pneumoniae* genome. Products from the present invention can be used in  
CC diagnosis kits and assays, and pharmaceutical compositions and vaccines  
CC for *S. pneumoniae*

XX  
SQ Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 0 U; 3 Other;

Query Match 92.4%; Score 19.4; DB 2; Length 8195;  
Best Local Similarity 95.2%; Pred. No. 17;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21  
Db 1384 TGCATGTATGGGTTGCTTCC 1364

## RESULT 44

ABSS56454\_09  
Continuation (10 of 22) of ABSS56454 from base 900001 (Streptococcus pneumoniae type 4  
WP Sequence split into 22 fragments LOCUS ABSS56454 Accession ABSS56454

WP	Fragment Name	Begin	End
WP	ABSS56454_00	1	110000
WP	ABSS56454_01	100001	210000
WP	ABSS56454_02	200001	310000
WP	ABSS56454_03	300001	410000
WP	ABSS56454_04	400001	510000
WP	ABSS56454_05	500001	610000
WP	ABSS56454_06	600001	710000
WP	ABSS56454_07	700001	810000
WP	ABSS56454_08	800001	910000
WP	ABSS56454_09	900001	1010000
WP	ABSS56454_10	1000001	1110000
WP	ABSS56454_11	1100001	1210000
WP	ABSS56454_12	1200001	1310000
WP	ABSS56454_13	1300001	1410000
WP	ABSS56454_14	1400001	1510000
WP	ABSS56454_15	1500001	1610000
WP	ABSS56454_16	1600001	1710000
WP	ABSS56454_17	1700001	1810000
WP	ABSS56454_18	1800001	1910000
WP	ABSS56454_19	1900001	2010000
WP	ABSS56454_20	2000001	2110000
WP	ABSS56454_21	2100001	2162598

Query Match 92.4%; Score 19.4; DB 10; Length 110000;  
Best Local Similarity 95.2%; Pred. No. 23;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21  
Db 108618 TGCATGTATGGGTTGCTTCC 108638

## RESULT 45

ABSS56454\_10  
Continuation (11 of 22) of ABSS56454 from base 1000001 (Streptococcus pneumoniae type 4  
WP Sequence split into 22 fragments LOCUS ABSS56454 Accession ABSS56454

WP	Fragment Name	Begin	End
WP	ABSS56454_00	1	110000
WP	ABSS56454_01	100001	210000
WP	ABSS56454_02	200001	310000
WP	ABSS56454_03	300001	410000
WP	ABSS56454_04	400001	510000
WP	ABSS56454_05	500001	610000
WP	ABSS56454_06	600001	710000
WP	ABSS56454_07	700001	810000
WP	ABSS56454_08	800001	910000
WP	ABSS56454_09	900001	1010000
WP	ABSS56454_10	1000001	1110000
WP	ABSS56454_11	1100001	1210000
WP	ABSS56454_12	1200001	1310000

WP	ABSS56454_13	1300001	1410000
WP	ABSS56454_14	1400001	1510000
WP	ABSS56454_15	1500001	1610000
WP	ABSS56454_16	1600001	1710000
WP	ABSS56454_17	1700001	1810000
WP	ABSS56454_18	1800001	1910000
WP	ABSS56454_19	1900001	2010000
WP	ABSS56454_20	2000001	2110000
WP	ABSS56454_21	2100001	2162598

Query Match 92.4%; Score 19.4; DB 10; Length 110000;  
Best Local Similarity 95.2%; Pred. No. 23;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21  
Db 8618 TGCATGTATGGGTTGCTTCC 8638

## RESULT 46

ACH68181  
ID ACH68181 standard; DNA; 553 BP.

XX	ACH68181;
XX	29-JUL-2004 (first entry)
XX	Human genome derived single exon probe #1376.
XX	Human; probe; ss; Gene expression; single exon probe; microarray;
XX	alternative splicing event; genomic alteration.
XX	Homo sapiens.
XX	US2003194704-A1.
XX	16-OCT-2003.
XX	03-APR-2002; 2002US-00029386.
XX	03-APR-2002; 2002US-00029386.
XX	(PENN/) PENN S G.
XX	(RANK/) RANK D R.
XX	(HANZ/) HANZEL D K.

Penn SG, Rank DR, Hanzel DK;  
WPI; 2004-119264/12.

XX New human genome-derived single exon nucleic acid probes useful for human  
XX gene expression analysis, for identifying or characterizing alternative  
XX splicing events, for assessing genomic alterations or as tools for  
XX surveying tissues.

XX Claim 15; SEQ ID NO 1376; 80pp; English.

XX The invention relates to a nucleic acid probe for measuring human gene  
XX expression, comprising any of the 27,400 fully defined nucleotide  
XX sequences in the specification, or their complements or fragments, and  
XX encoding at least 8 amino acids of any of the 688 amino acid sequences  
XX fully defined in the specification. The probe is a single exon probe that  
XX hybridizes under high stringency conditions to a nucleic acid molecule  
XX expressed in human cells or tissues. Also included are a spatially-  
XX addressable set of single exon nucleic acid probes for measuring human  
XX gene expression (comprising a plurality of single exon nucleic acid  
XX probes cited above, where each of the plurality of probes is separately  
XX and addressably isolatable or amplifiable from the plurality), a single  
XX exon microarray for measuring human gene expression, a method of  
XX measuring human gene expression, a vector comprising the single exon  
XX probe cited above, an ORF-encoded peptide comprising at least 8  
XX contiguous amino acids of any of the above-mentioned amino acid  
XX sequences (optionally with conservative amino acid substitutions), an

CC isolated antibody that binds specifically to a peptide cited above,  
CC methods of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subcription, and a method of providing  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterising  
CC alternative splicing events, in detecting and characterising gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe of the invention. Note: The present sequence for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030194704  
XX  
SQ Sequence 553 BP; 125 A; 160 C; 151 G; 117 T; 0 U; 0 Other;

Query Match 87.6%; Score 18.4; DB 12; Length 553;  
Best Local Similarity 95.0%; Pred. No. 39;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCATGTATGGGTATCTTC 20  
||||||| |||||||  
Db 213 TGCATGTATGATGTTATCTTC 232

RESULT 47  
ADBI1789/c  
ID ADBI1789 standard; DNA; 1830 BP.  
XX  
AC ADBI1789;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Alloicoccus otitis antigenic protein encoding DNA SEQ ID NO:5273.  
XX  
KW Alloicoccus otitis; antigenic protein; immunogenic; immunisation;  
KW gene therapy; Gram-positive bacterium; infection; gene; ds.  
XX  
OS Alloicoccus otitis.  
XX  
PN WO2003048304-A2.  
XX  
PD 12-JUN-2003.  
XX  
PF 25-NOV-2002; 2002WO-US036123.  
XX  
PR 29-NOV-2001; 2001US-0333777P.  
PR 18-NOV-2002; 2002US-0426742P.  
XX  
PA (AMHP ) WYETH HOLDINGS CORP.  
XX  
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;  
XX WPI; 2003-505284/47.  
XX P-PSDB; ADBI1786.  
XX  
XX New Alloicoccus otitis polynucleotides and polypeptides, useful for  
XX treating and diagnosing diseases, drug screening assays and monitoring of  
XX effects during drug clinical trials.  
XX  
XX Claim 7; SEQ ID NO 5273; 1019pp; English.

XX The present invention describes an isolated polynucleotide (I) of  
CC Alloicoccus otitis genomic DNA, which encodes an antigenic protein.  
CC Alloicoccus otitis is a Gram-positive bacterium. Also described: (1)  
CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
CC expression vector comprising the novel isolated polynucleotide (I), its

CC complement, degenerate variant or fragment; (3) a genetically engineered  
CC host cell, transfected, transformed or infected with the vector of (2);  
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
CC composition comprising the polypeptide, its complement, biological  
CC equivalent or fragment, or the polynucleotide that is comprised in the  
CC expression vector; (6) a pharmaceutical composition comprising the  
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
CC immunising against Alloicoccus otitis by administering to a host the  
CC immunogenic composition; (9) detecting and/or identifying Alloicoccus  
CC otitis in the biological sample; (10) a kit comprising a container  
CC containing the novel polynucleotide, its degenerate variant or fragment,  
CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
CC genetically engineered host cell under conditions suitable to produce the  
CC polypeptide from the culture. (I) can be used in gene therapy. The  
CC polynucleotides, polypeptides, antibodies and compositions of the present  
CC invention can be used for treating and diagnosing diseases, drug  
CC screening assays and monitoring of effects during drug clinical trials.  
CC The polynucleotides are useful for expressing and detecting Alloicoccus  
CC otitis. The present sequence encodes an Alloicoccus otitis antigen  
CC protein from the present invention.

SQ Sequence 1830 BP; 565 A; 389 C; 453 G; 423 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 9; Length 1830;  
Best Local Similarity 90.5%; Pred. No. 89;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCATGTATGGGTATCTTC 21  
||||||| |||||||  
Db 485 TGCATGTAGGGTGTCTTC 465

RESULT 48  
ADBI1785/c  
ID ADBI1785 standard; DNA; 1830 BP.  
XX  
AC ADBI1785;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Alloicoccus otitis antigenic protein encoding DNA SEQ ID NO:5277.  
XX  
KW Alloicoccus otitis; antigenic protein; immunogenic; immunisation;  
KW gene therapy; Gram-positive bacterium; infection; gene; ds.  
XX  
OS Alloicoccus otitis.  
XX  
PN WO2003048304-A2.  
XX  
PD 12-JUN-2003.  
XX  
PF 25-NOV-2002; 2002WO-US036123.  
XX  
PR 29-NOV-2001; 2001US-0333777P.  
PR 18-NOV-2002; 2002US-0426742P.  
XX  
PA (AMHP ) WYETH HOLDINGS CORP.  
XX  
XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;  
XX WPI; 2003-505284/47.  
XX P-PSDB; ADBI1782.  
XX  
XX New Alloicoccus otitis polynucleotides and polypeptides, useful for  
XX treating and diagnosing diseases, drug screening assays and monitoring of  
XX effects during drug clinical trials.  
XX  
XX Claim 7; SEQ ID NO 5277; 1019pp; English.

XX The present invention describes an isolated polynucleotide (I) of  
CC Alloicoccus otitis genomic DNA, which encodes an antigenic protein.  
CC Alloicoccus otitis is a Gram-positive bacterium. Also described: (1)  
CC an isolated polypeptide that is encoded by the polynucleotide (I); (2) an  
CC expression vector comprising the novel isolated polynucleotide (I), its

CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an  
CC expression vector comprising the novel isolated polynucleotide (1); its  
CC complement, degenerate variant or fragment; (3) a genetically engineered  
CC host cell, transfected, transformed or infected with the vector of (2);  
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
CC composition comprising the polypeptide, its complement, biological  
CC equivalent or fragment, or the polynucleotide that is comprised in the  
CC expression vector; (6) a pharmaceutical composition comprising the  
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
CC immunising against *Alloiooccus* otitis by administering to a host the  
CC immunogenic composition; (9) detecting and/or identifying *Alloiooccus*  
CC otitis in the biological sample; (10) a kit comprising a container  
CC containing the novel polynucleotide, its degenerate variant or fragment,  
CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
CC genetically engineered host cell under conditions suitable to produce the  
CC polypeptide from the culture. (1) can be used in gene therapy. The  
CC polynucleotides, polypeptides, antibodies and compositions of the present  
CC invention can be used for treating and diagnosing diseases, drug  
CC screening assays and monitoring of effects during drug clinical trials.  
CC The polynucleotides are useful for expressing and detecting *Alloiooccus*  
CC otitis. The present sequence encodes an *Alloiooccus* otitis antigen  
CC protein from the present invention.

XX  
SQ Sequence 1830 BP; 565 A; 389 C; 453 G; 423 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 9; Length 1830;  
Best Local Similarity 90.5%; Pred. No. 89;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTATCTTCC 21  
||||| ||||| ||||| |||||  
Db 485 TGCATGATGGGTATCTTCC 465

RESULT 49  
ADB11787/C  
ID ADB11787 standard; DNA; 1830 BP.  
XX ADB11787;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX *Alloiooccus* otitis antigenic protein encoding DNA SEQ ID NO:5275.

XX *Alloiooccus* otitis; antigenic protein; immunogenic; immunisation;  
KW gene therapy; Gram-positive bacterium; infection; gene; ds.  
XX *Alloiooccus* otitis.

OS *Alloiooccus* otitis.

XX WO2003048304-A2.

XX 12-JUN-2003.

XX 25-NOV-2002; 2002WO-US036123.

XX 29-NOV-2001; 2001US-0333777P.

XX 18-NOV-2002; 2002US-0426742P.

XX (AMHP ) WYETH HOLDINGS CORP.

XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

XX P-PSDB; ADB11784.

XX New *Alloiooccus* otitis polynucleotides and polypeptides, useful for  
PT treating and diagnosing diseases, drug screening assays and monitoring of  
PT effects during drug clinical trials.

XX Claim 7; SEQ ID NO 5275; 1019pp; English.

XX The present invention describes an isolated polynucleotide (1) of

CC *Alloiooccus* otitis genomic DNA, which encodes an antigenic protein.  
CC *Alloiooccus* otitis is a Gram-positive bacterium. Also described: (1)  
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an  
CC expression vector comprising the novel isolated polynucleotide (1); its  
CC complement, degenerate variant or fragment; (3) a genetically engineered  
CC host cell, transfected, transformed or infected with the vector of (2);  
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
CC composition comprising the polypeptide, its complement, biological  
CC equivalent or fragment, or the polynucleotide that is comprised in the  
CC expression vector; (6) a pharmaceutical composition comprising the  
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
CC immunising against *Alloiooccus* otitis by administering to a host the  
CC immunogenic composition; (9) detecting and/or identifying *Alloiooccus*  
CC otitis in the biological sample; (10) a kit comprising a container  
CC containing the novel polynucleotide, its degenerate variant or fragment,  
CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
CC genetically engineered host cell under conditions suitable to produce the  
CC polypeptide from the culture. (1) can be used in gene therapy. The  
CC polynucleotides, polypeptides, antibodies and compositions of the present  
CC invention can be used for treating and diagnosing diseases, drug  
CC screening assays and monitoring of effects during drug clinical trials.  
CC The polynucleotides are useful for expressing and detecting *Alloiooccus*  
CC otitis. The present sequence encodes an *Alloiooccus* otitis antigen  
CC protein from the present invention.

XX  
SQ Sequence 1830 BP; 565 A; 389 C; 453 G; 423 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 9; Length 1830;  
Best Local Similarity 90.5%; Pred. No. 89;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTATCTTCC 21  
||||| ||||| ||||| |||||  
Db 485 TGCATGATGGGTATCTTCC 465

RESULT 50  
ADB11783/C  
ID ADB11783 standard; DNA; 1830 BP.  
XX ADB11783;  
XX  
XX 20-NOV-2003 (first entry)  
XX  
XX *Alloiooccus* otitis antigenic protein encoding DNA SEQ ID NO:5279.

XX *Alloiooccus* otitis; antigenic protein; immunogenic; immunisation;  
KW gene therapy; Gram-positive bacterium; infection; gene; ds.

XX *Alloiooccus* otitis.

XX WO2003048304-A2.

XX 12-JUN-2003.

XX 25-NOV-2002; 2002WO-US036123.

XX 29-NOV-2001; 2001US-0333777P.

XX 18-NOV-2002; 2002US-0426742P.

XX (AMHP ) WYETH HOLDINGS CORP.

XX Fletcher LD, McMichael JC, Russell DP, Zagursky RJ;

XX WPI; 2003-505284/47.

XX P-PSDB; ADB11780.

XX New *Alloiooccus* otitis polynucleotides and polypeptides, useful for  
PT treating and diagnosing diseases, drug screening assays and monitoring of  
PT effects during drug clinical trials.

XX Claim 7; SEQ ID NO 5279; 1019pp; English.

XX The present invention describes an isolated polynucleotide (1) of  
CC Alloococcus otitidis genomic DNA, which encodes an antigenic protein.  
CC an isolated polypeptide that is encoded by the polynucleotide (1); (2) an  
CC expression vector comprising the novel isolated polynucleotide (1), its  
CC complement, degenerate variant or fragment; (3) a genetically engineered  
CC host cell, transfected, transformed or infected with the vector of (2);  
CC (4) an antibody specific for the polypeptide of (1); (5) an immunogenic  
CC composition comprising the polypeptide, its complement, biological  
CC equivalent or fragment, or the polynucleotide that is comprised in the  
CC expression vector; (6) a pharmaceutical composition comprising the  
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array  
CC of the polypeptides of (1), their biological equivalent or fragment; (8)  
CC immunising against Alloococcus otitidis by administering to a host the  
CC antigenic composition; (9) detecting and/or identifying Alloococcus  
CC otitidis in the biological sample; (10) a kit comprising a container  
CC containing the novel polynucleotide, its degenerate variant or fragment,  
CC or the antibody of (4); and (11) producing a polypeptide by culturing the  
CC genetically engineered host cell under conditions suitable to produce the  
CC polypeptide from the culture. (1) can be used in gene therapy. The  
CC polynucleotides, polypeptides, antibodies and compositions of the present  
CC invention can be used for treating and diagnosing diseases, drug  
CC screening assays and monitoring of effects during drug clinical trials.  
CC The polynucleotides are useful for expressing and detecting Alloococcus  
CC otitidis. The present sequence encodes an Alloococcus otitidis antigen  
CC protein from the present invention.

XX SQ Sequence 1830 BP; 565 A; 389 C; 453 G; 423 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 9; Length 1830;  
Best Local Similarity 90.5%; Pred. No. 89;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 485 TGCATGTAGGGTGTCTTCC 465

RESULT 51  
ID ADR32245/c  
XX ADR32245 standard; cDNA; 2914 BP.  
AC ADR32245;  
XX 07-OCT-2004 (first entry)  
XX Human tumour suppressor BNO230 cDNA, SEQ ID NO:15.  
XX Tumour suppressor; cancer; breast; chromosome 16q24.3;  
KW loss of heterozygosity; LOH; transgenic animal; drug screening;  
KW diagnosis; prognosis; cytostatic; gene therapy; human; BNO230; FLJ31875;  
KW zinc finger domain; transcriptional regulation; gene; ss.

XX Homo sapiens.  
XX Key Location/Qualifiers  
FH 333..1652  
FT /\*tag= a  
FT /product= "Tumour suppressor protein BNO230"

XX WO200264780-A1.  
XX 22-AUG-2002.  
XX 12-FEB-2002; 2002WO-AU000138.  
XX 12-FEB-2001; 2001AU-00003052.  
PR 12-FEB-2001; 2001AU-00003053.  
PR 12-FEB-2001; 2001AU-00003054.  
XX (BION-) BIONOMICS LTD.

XX PA

PI Callen DF, Powell JA, Kremmidiotis G, Gardner AB, Whitmore SA;  
XX WPI; 2002-667005/71.  
DR P-PSDB; ADR32246.

XX Tumor suppressor genes at chromosome 16q24.3, useful for diagnosing,  
PT inhibiting or treating cancer or establishing the prognosis of a patient  
PT diagnosed with cancer, preferably breast cancer.

XX Claim 11; SEQ ID NO 15; 129pp; English.

XX The invention relates to 13 tumour suppressor genes involved in breast  
CC cancer, and their encoded proteins (ADR32231-ADR32256). The tumour  
CC suppressor proteins are PRDM7 (PR domain containing 7 protein), CDT1  
CC (CDT1 DNA replication factor), CHMP1 (charged multivesicular body protein),  
CC 1/chromatin modifying protein 1), BANP (BTG3 associated nuclear protein),  
CC BNO224, BNO36, BNO34, BNO208, BNO230 and BNO44. The genes encoding these  
CC proteins are located on chromosome 16q24.3, a region exhibiting loss of  
CC heterozygosity (LOH) in breast cancer and also in other carcinomas such  
CC as prostate cancers. The invention also relates to vectors and host cells  
CC comprising a tumour suppressor gene of the invention; antibodies against  
CC the tumour suppressor proteins; microarrays comprising fragments of the  
CC tumour suppressor genes; mutant forms of the tumour suppressor genes and  
CC proteins; methods of diagnosing, treating or inhibiting breast cancer; a  
CC genetically modified non-human animal in which a tumour suppressor gene  
CC is inactivated by knockout deletion; methods of screening for drugs which  
CC restore tumour suppressor activity; and pharmaceutical compositions  
CC comprising such a drug or a tumour suppressor gene, protein or antibody.  
CC The tumour suppressor genes and proteins of the invention may be used for  
CC diagnosing, inhibiting or treating cancer or for establishing the  
CC prognosis of a patient diagnosed with cancer, particularly breast cancer.  
CC The present sequence represents a cDNA encoding the tumour suppressor  
CC protein BNO230 (also known as FLJ31875). The BNO230 protein contains zinc  
CC finger domains, suggesting a role in DNA binding and/or protein complex  
CC interactions that regulate transcription.

XX SQ Sequence 2914 BP; 821 A; 700 C; 728 G; 665 T; 0 U; 0 Other;

Query Match 84.8%; Score 17.8; DB 7; Length 2914;  
Best Local Similarity 90.5%; Pred. No. 94;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 1517 TGCATGTATGGGTATTATCC 1497

RESULT 52  
ID ADA53346/c  
XX ADA53346 standard; cDNA; 2914 BP.

AC ADA53346;

XX 20-NOV-2003 (first entry)

XX Human coding sequence, SEQ ID 914.

XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;  
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;  
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.

XX Homo sapiens.

XX EPI293569-A2.

XX 19-MAR-2003.

XX 21-MAR-2002; 2002EP-00006586.

XX 14-SEP-2001; 2001JP-0028381.

XX 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.



PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
XX  
PI Isogai T, Sugiyana T, Otsuki T, Wakamatsu A, Sato H, Ishii S;  
PI Yamamoto J, Isono Y, Hlo Y, Otsuka K, Nagai K, Irie R, Tamechika I;  
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;  
XX  
DR WPI; 2003-395539/38.  
DR P-PSDB; ADA54985.  
XX  
XX New polynucleotides encoding full-length polypeptides, e.g. secretory  
PT and/or membrane proteins, useful for developing medicines for diseases in  
PT which the gene is involved, or as target molecules for gene therapy.  
XX  
XX Claim 1; SEQ ID NO 914; 205pp; English.  
XX  
CC The present invention relates to novel human secretory or membrane  
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-  
CC ADA54071). The coding sequences are useful in the gene therapy of  
CC diseases caused by abnormalities of the proteins, e.g. cancer,  
CC inflammatory diseases, osteoporosis or neurological disease.  
XX  
XX Sequence 2914 BP; 821 A; 700 C; 728 G; 665 T; 0 U; 0 Other;  
SQ  
  
Query Match 84.8%; Score 17.8; DB 10; Length 2914;  
Best Local Similarity 90.5%; Pred. No. 94;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
DB 1517 TGCATGTATGGGTATCTTCC 1497  
  
RESULT 53  
ADBI2064\_13  
Continuation (14 of 18) of ADBI2064 from base 1300001 (Allootococcus otitis entire genome  
WP Sequence split into 18 fragments LOGUS ADBI2064 Accession Adbi2064  
WP Fragment Name Begin End  
WP ADBI2064\_00 1 110000  
WP ADBI2064\_01 100001 210000  
WP ADBI2064\_02 200001 310000  
WP ADBI2064\_03 300001 410000  
WP ADBI2064\_04 400001 510000  
WP ADBI2064\_05 500001 610000  
WP ADBI2064\_06 600001 710000  
WP ADBI2064\_07 700001 810000  
WP ADBI2064\_08 800001 910000  
WP ADBI2064\_09 900001 1010000  
WP ADBI2064\_10 1000001 1110000  
WP ADBI2064\_11 1100001 1210000  
WP ADBI2064\_12 1200001 1310000  
WP ADBI2064\_13 1300001 1410000  
WP ADBI2064\_14 1400001 1510000  
WP ADBI2064\_15 1500001 1610000  
WP ADBI2064\_16 1600001 1710000  
WP ADBI2064\_17 1700001 1754382  
  
Query Match 84.8%; Score 17.8; DB 9; Length 110000;  
Best Local Similarity 90.5%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 1 TGCATGTATGGGTATCTTCC 21  
|||||  
DB 94003 TGCATGTATGGGTATCTTCC 94023  
  
RESULT 54  
ABK74374  
ID ABK74374 standard; DNA; 412 BP.  
XX  
AC ABK74374;  
XX  
DT 13-AUG-2002 (first entry)  
XX

DE Bacillus licheniformis genomic sequence tag (GST) #1665.  
XX  
KW Differential gene expression; genomic sequenced tag; GST;  
KW altered culture condition; environmental stress;  
KW physiological provocation; ds.  
XX  
OS Bacillus licheniformis.  
XX  
PN WO200229113-A2.  
XX  
PD 11-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US031437.  
XX  
PR 06-OCT-2000; 2000US-00680598.  
PR 27-MAR-2001; 2001US-0279526P.  
XX  
PA (NOVO ) NOVOZYMES BIOTECH INC.  
PA (NOVO ) NOVOZYMES AS.  
XX  
PI Berka R, Clausen IG;  
XX  
DR WPI; 2002-416684/44.  
XX  
XX Monitoring differential expression of several genes in first Bacillus  
PT cell relative to expression of same genes in one or more second Bacillus  
PT cells, by using substrate containing Bacillus genomic sequenced tag  
PT array.  
XX  
PS Claim 4; SEQ ID NO 1665; 200pp; English.  
XX  
CC The invention describes a method of monitoring differential expression of  
CC genes in a first Bacillus cell relative to expression of the genes in  
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes  
CC isolated from Bacillus cells to a substrate containing array of Bacillus  
CC genomic sequenced tags (GST), examining the array, and determining  
CC relative gene expression by an observed hybridisation reporter signal of  
CC a spot in the array. The method is useful for measuring the expression of  
CC genes in a first Bacillus cell relative to expression of the same genes  
CC in one or more second Bacillus cells. The method is useful for monitoring  
CC global expression of several genes from a Bacillus cell, discovering new  
CC genes, identifying possible functions of unknown open reading frames and  
CC monitoring gene copy number variation and stability. Monitoring changes  
CC in expression of genes may be used to provide a representation of the way  
CC in which Bacillus cells adapt to changes in culture conditions, follow  
CC environmental stress or other physiological provocation. Extensive follow  
CC -up characterisation is unnecessary, when one spot on an array equals one  
CC gene or one open reading frame, since sequence information is available.  
CC This sequence represents a genomic sequence tag (GST) used in the method  
CC of the invention. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 412 BP; 137 A; 74 C; 104 G; 97 T; 0 U; 0 Other;  
  
Query Match 80.0%; Score 16.8; DB 6; Length 412;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 GCATGTATGGGTATCTTCC 21  
|||||  
DB 78 GCATGTATGGGTATCTGCC 97  
  
RESULT 55  
AAX30475/c  
ID AAX30475 standard; DNA; 1464 BP.  
XX  
AC AAX30475;  
XX  
DT 08-JUN-1999 (first entry)  
XX  
DE H. pylori secreted protein ORF hp6p10606\_23493756\_c1\_21.



XX Vaccine; probe; diagnostic; ORF; cell envelope protein; secreted protein;  
 KW cellular protein; ds.

XX Helicobacter pylori.

XX WO9818323-A1.

XX 07-MAY-1998.

XX 28-OCT-1997; 97WO-US019575.

XX 28-OCT-1996; 96US-00739150.

XX 06-DEC-1986; 96US-00759739.

XX 14-JUL-1997; 97US-00891928.

XX (ASTR ) ASTRA AB.

XX Smith D, Alm RA;

XX WPI; 1998-271811/24.

XX P-PSDB; AAY11008.

XX Helicobacter pylori nucleic acids and proteins - used to develop products  
 for the detection, prevention and treatment of H. pylori infections.

XX Claim 3, 4; Page 137; 279pp; English.

XX Recombinant or substantially pure preparations of H. pylori polypeptides  
 are disclosed, together with the nucleic acids encoding them. In all, 73  
 ORFs are shown. The proteins are variously cell envelope proteins,  
 secreted proteins or other cellular proteins. Vaccines containing the  
 nucleic acids or proteins are claimed, as are probes containing at least  
 8 nucleotides from the nucleic acid sequences. The vaccines are useful  
 for treating or reducing the risk of H. pylori infections, and the probes  
 can be used diagnostically for detecting the presence of Helicobacter in  
 a sample. The products are also of use in screening for compounds having  
 the ability to interfere with the H. pylori life cycle or to inhibit H.  
 pylori infection

SQ Sequence 1464 BP; 545 A; 273 C; 276 G; 370 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 2; Length 1464;

Best Local Similarity 90.0%; Pred. No. 2.7e+02; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 2;

Qy 1 TGCATGTATGGTTTATCTTC 20

Db 1370 TGCATGTATGGTTTCTTC 1351

RESULT 56

AAAT68140/c

ID AAAT68140 standard; DNA; 1554 BP.

XX AAAT68140;

XX 18-JUL-1997 (first entry)

DE H. pylori inner membrane protein ORF 13ep12003orf20.

XX Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis; ds.

XX Helicobacter pylori.

XX Key Location/Qualifiers

FT 1. .1554

FT /\*tag= a

FT /note= "no stop codon given"

PN WO9640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96WO-US009122.

XX 07-JUN-1995; 96US-00487032.

XX 01-APR-1996; 96US-00630405.

XX (ASTR ) ASTRA AB.

XX Smith D, Berglinth OT, Mellgaard BL;

XX WPI; 1997-052306/05.

XX P-PSDB; AAW20887.

XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 PT useful for vaccines to treat or prevent H. pylori infection, and to  
 PT detect Helicobacter.

XX Claim 1; Page 924; 1481pp; English.

XX This sequence encodes a H. pylori inner membrane protein. The protein may  
 be used in a vaccine to prevent or treat H. pylori infection or to  
 identify H. pylori polypeptide binding compounds, useful as potential H.  
 pylori life cycle activators or inhibitors. The genomic sequence of H.  
 pylori (ATCC 55679) was determined from overlapping contigs generated by  
 mechanically shearing the bacterial DNA. The sequences were analysed for  
 ORF of at least 180 nucleotides, and the predicted coding regions defined  
 by computer evaluation. To identify likely H. pylori antigens for vaccine  
 development, the amino acid sequences predicted from various ORF were  
 analysed for significant homology to other known or exported membrane  
 proteins. Having identified and determined the sequences of interest,  
 particular regions can be isolated from H. pylori by PCR amplification  
 for recombinant polypeptide production, e.g. in E. coli hosts

SQ Sequence 1554 BP; 571 A; 281 C; 290 G; 412 T; 0 U; 0 Other;

Query Match 80.0%; Score 16.8; DB 2; Length 1554;

Best Local Similarity 90.0%; Pred. No. 2.8e+02;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTTATCTTC 20

Db 1463 TGCATGTATGGTTTCTTC 1444

RESULT 57

ADE91329/c

ID ADE91329 standard; DNA; 2547 BP.

XX ADE91329;

XX 12-FEB-2004 (first entry)

DE Maize glutamine synthetase (GS1-2) promoter.

XX Maize; cytoplasmic glutamine synthetase promoter; GS1-2 promoter;  
 KW transgene expression; plant breeding; corn; transgenic; ds.

XX Zea mays.

XX US2003140364-A1.

XX 24-JUL-2003.

XX 20-NOV-2001; 2001US-00989739.

XX 20-NOV-2001; 2001US-00989739.

XX (HINC/) HINCHEY B.

XX (SONG/) SONG H.

```
PI Hinchey B, Song H;
XX WPI; 2004-020397/02.
XX
XX Novel isolated nucleic acid sequence comprising cytoplasmic glutamine
PT synthetase promoter useful for improving feed or food value, improving
PT processing of corn and products resulting from processing.
XX
XX Claim 11; Fig 2; 62pp; English.
XX
XX The present invention relates to the isolation of maize cytoplasmic
CC glutamine synthetase (GS1-2) promoter. Also disclosed are methods for the
CC expression of transgenes in plants, stably transformed with a selected
CC DNA comprising a maize GS1-2 promoter. The GS1-2 promoter and the methods
CC are useful for plant breeding. The transgenic plant is as a
CC monocotyledonous plant (wheat, maize, rice, oat, barley, turfgrass,
CC sorghum, millet and sugarcane) or dicotyledonous plant (tobacco, tomato,
CC potato, soybean, cotton, canola, alfalfa, sunflower, and cotton).
CC Preferably, the monocotyledonous or dicotyledonous plant is maize or
CC soybean plant, respectively. The promoter is useful for directing the
CC expression of a selected coding region which encodes a particular protein
CC or polypeptide product, improving feed or food value, improving
CC processing of corn and for improving the value of the product resulting
CC from the processing. The promoter is efficient in expressing transgenes
CC in plants. The present sequence represents maize GS1-2 promoter.
XX
XX Sequence 2547 BP; 691 A; 602 C; 574 G; 680 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16.8; DB 12; Length 2547;
Best Local Similarity 90.0%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTATCTTC 20
DB 1824 TGCATGCTGGGCTATCTTC 1805
RESULT 59
ABV72307/c
ID ABV72307 standard; DNA; 2953 BP.
AC
XX ABV72307;
XX
XX 13-MAY-2003 (first entry)
XX
XX Nucleotide sequence of fructosyl amino acid oxidase.
DE
XX Fructosyl amino acid oxidase; enzyme; gene; ss.
KW
XX Aspergillus oryzae.
OS
XX
FH Key Location/Qualifiers
FT CDS 1038..2551
FT /*tag= a
FT /product= "fructosyl amino acid oxidase"
FT /note= "contains introns"
FT exon 1038..1086
FT /*tag= b
FT /*number= 1
FT intron 1087..1224
FT /*tag= c
FT /*number= 1
FT exon 1225..2213
FT /*tag= d
FT intron 2214..2269
FT /*tag= e
FT /*number= 2
FT exon 2270..2551
FT /*tag= f
FT /*number= 3
XX JP2002218982-A.
PI Hinchey B, Song H;
XX WPI; 2004-020397/02.
XX
XX Novel isolated nucleic acid sequence comprising cytoplasmic glutamine
PT synthetase promoter useful for improving feed or food value, improving
PT processing of corn and products resulting from processing.
XX
XX Claim 11; Fig 2; 62pp; English.
XX
XX The present invention relates to the isolation of maize cytoplasmic
CC glutamine synthetase (GS1-2) promoter. Also disclosed are methods for the
CC expression of transgenes in plants, stably transformed with a selected
CC DNA comprising a maize GS1-2 promoter. The GS1-2 promoter and the methods
CC are useful for plant breeding. The transgenic plant is as a
CC monocotyledonous plant (wheat, maize, rice, oat, barley, turfgrass,
CC sorghum, millet and sugarcane) or dicotyledonous plant (tobacco, tomato,
CC potato, soybean, cotton, canola, alfalfa, sunflower, and cotton).
CC Preferably, the monocotyledonous or dicotyledonous plant is maize or
CC soybean plant, respectively. The promoter is useful for directing the
CC expression of a selected coding region which encodes a particular protein
CC or polypeptide product, improving feed or food value, improving
CC processing of corn and for improving the value of the product resulting
CC from the processing. The promoter is efficient in expressing transgenes
CC in plants. The present sequence represents maize GS1-2 promoter.
XX
XX Sequence 2547 BP; 691 A; 602 C; 574 G; 680 T; 0 U; 0 Other;
SQ
Query Match 80.0%; Score 16.8; DB 6; Length 2953;
Best Local Similarity 90.0%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGCATGTATGGGTATCTTC 20
DB 1441 TGCATGTATCGGTATCTTC 1422
RESULT 59
ADZ12700/c
ID ADZ12700 standard; DNA; 48179 BP.
AC
XX ADZ12700;
XX
XX 16-JUN-2005 (first entry)
XX
XX Murine cancer-associated genomic DNA #18.
DE
XX
XX Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
KW cytostatic; gene; ds.
XX
XX Mus sp.
OS
XX
XX WO2005031001-A2.
PN
XX
XX 07-APR-2005.
XX
XX 23-SEP-2004; 2004WO-US031617.
XX
XX 23-SEP-2003; 2003US-00669920.
XX (CHIR ) CHIRON CORP.
XX
XX Morris DW, Melandro MS;
PI
XX WPI; 2005-273395/28.
XX
XX Nucleic acid array useful for detecting cancer associated nucleic acid,
PT comprises two or more nucleic acid probes.
XX
XX Disclosure; SEQ ID NO 220; 198pp; English.
XX
XX The invention relates to a nucleic acid array for detecting a cancer
CC associated (CA) nucleic acid, comprising two or more nucleic acid probes.
CC The invention also relates to a peptide array comprising two or more
CC isolated polypeptides encoded by a CA nucleic acid sequence, a compound
CC that binds to a polypeptide, an isolated antibody or its fragment which
CC binds to a polypeptide, which is prepared by immunizing a host animal
CC with a composition comprising the polypeptide or its antigen binding
CC fragment and collecting cells from the host expressing antibodies against
```

CC the antigen or its antigen binding fragment, a composition comprising the  
CC antibody and a carrier, a method of screening for anticancer activity, a  
CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a  
CC method of treating cancer and a method of inhibiting expression of a CA  
CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA  
CC nucleic acids. The antibody is useful for detecting the presence or  
CC absence of cancer cells in an individual which involves contacting cells  
CC from the individual with the antibody and detecting a complex of a CA  
CC protein from the cancer cells and the antibody, where the detection of  
CC the complex correlates with the presence of cancer cells in the  
CC individual. The composition is useful for inhibiting growth of cancer  
CC cells in an individual or for delivering a therapeutic agent to cancer  
CC cells in an individual. The invention is also useful for diagnosing  
CC cancer, for treating cancer and for inhibiting expression of a CA gene in  
CC a cell. This sequence represents murine cancer-associated genomic DNA of  
CC the invention.

XX  
SQ Sequence 48179 BP; 13764 A; 10405 C; 10337 G; 13482 T; 0 U; 191 Other;

Query Match 80.0%; Score 16.8; DB 14; Length 48179;  
Best Local Similarity 90.0%; Pred. No. 4.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTCC 21  
||||| ||||| ||||| |||||  
Db 34371 GCATGTGTGGTTATCTTCC 34352

RESULT 60  
ADA02873/c  
ID ADA02873 standard; DNA; 51281 BP.  
AC ADA02873;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Mouse Dpt carcinoma associated gene, SEQ ID NO:1391.  
XX  
KW Mouse; murine; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ds.  
XX  
OS Mus sp.

XX WO2003057146-A2.

XX 17-JUL-2003.

XX 26-DEC-2002; 2002WO-US041414.

XX 26-DEC-2001; 2001US-00035832.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW;

XX WPI; 2003-587068/55.

XX New recombinant nucleic acid encoding carcinoma associated protein,  
XX useful for preparing compositions for treating carcinomas.

XX Claim 1; SEQ ID NO 1391; 245pp; English.

XX The invention relates to recombinant carcinoma associated (CA) nucleic  
XX acid sequences from mouse and human (ADA01482-ADA03094), and to  
XX recombinant carcinoma associated proteins (CAP) encoded by them. The  
XX invention also encompasses expression vectors and host cells comprising a  
XX CA nucleic acid, a polypeptide (especially an antibody) that specifically  
XX binds to the protein, and a biochip comprising CA nucleic acid or  
XX fragments thereof. The sequences of the invention were identified using  
XX oncogenic retroviruses, which insert into the genome of the host organism  
XX at random. Many of these do not carry transduced host oncogenes or  
XX pathogenic trans-acting viral genes, meaning that cancer incidence is a

CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed murine CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 51281 BP; 13608 A; 10106 C; 10141 G; 12908 T; 0 U; 4518 Other;

Query Match 80.0%; Score 16.8; DB 9; Length 51281;

Best Local Similarity 90.0%; Pred. No. 4.2e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTCC 21

||||| ||||| ||||| |||||  
Db 37474 GCATGTGTGGTTATCTTCC 37455

Search completed: January 27, 2006, 22:15:20  
Job time : 201.315 secs

***This Page Blank (uspto)***



```

/mol_type="mRNA"
/db_xref="taxon:338618"
/clone="ColPK16"
/tissue_type="mixed shoot and floral apical meristems,
flower buds, leaves and roots"
/lab_host="DH10B T1 (T1 and T5 phage resistance)"
/clone_lib="Aquilgia cDNA library"
/note="Vector: pCMV SPORT6.1; Site 1: EcoRI; Site 2: NotI;
F2, F3, and F4 lines of Aquilgia formosa X A. pubescens
were grown from seed in greenhouses at UC Santa Barbara.
From these plants three sets of tissue were collected: 1)
Small flower buds (<10 mm) and very young inflorescences
(71 & 29% by weight respectively), 2) Medium (7-20 mm) and
large (at or near anthesis) flower buds (65 & 35% by
weight respectively) and 3) Shoot apical meristems. A
fourth set of tissue was collected from plants of A.
formosa. These plants were grown from seed in sand at
approximately 1 month root tissue and leaf tissue of
various developmental stages were collected (84 & 16% by
weight respectively). Total RNA was extracted from each
set of tissue and pooled in the following proportions:
1.5X from sets 1 & 2, 1X from sets 3 & 4. From the pooled
total RNA, mRNA was extracted and enriched for full-length
messages and then normalized with proprietary methods by
Invitrogen."

```

## ORIGIN

```

Query Match      87.6%; Score 18.4; DB 8; Length 815;
Best Local Similarity 95.0%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 TGCATGTATGGGTATCTTC 20
        ||||| ||||| ||||| |||||
DB      278 TGCATGTACGGGTATCTTC 259

```

## RESULT 2

```

BX556085
LOCUS      BX556085 Glossina morsitans morsitans adult infected gut Glossina
DEFINITION morsitans morsitans cDNA clone Tse24g11_q1c, mRNA sequence.

```

```

ACCESSION  BX556085.1 GI:33380027
VERSION     BX556085
KEYWORDS   EST.
SOURCE     Glossina morsitans morsitans

```

## ORGANISM

```

Glossina morsitans morsitans
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Hippoboscidae; Glossinidae; Glossina.

```

## REFERENCE

```

1 (bases 1 to 443)
Lehane, M.J., Aksoy, S., Gibson, W., Kerhornou, A., Berriman, M.,
Hamilton, J., Soares, M.B., Bonaldo, M.F., Lehane, S. and Hall, N.
Adult midgut expressed sequence tags from the tsetse fly Glossina
morsitans morsitans and expression analysis of putative immune
response genes

```

```

JOURNAL     Genome Biol. 4 (10), R63 (2003)
PUBMED     14519198
COMMENT     Contact: Hall N

```

```

Pathogen Sequencing Unit
The Sanger Institute The Wellcome Trust Genome Campus
Hinxton, Cambridge, CB10 1SA, UK
Request for clones, please contact: Mike Lehane
Prof. M.J. Lehane
School of Biological Sciences,
University of Wales,
Bangor LL57 2UW

```

```

All clones with suffix q1c are reverse primer reads starting at 5'
end of the cDNA all p1c reads are from
the 3' end.

```

## FEATURES

```

source      Location/Qualifiers
1..443
/organism="Glossina morsitans morsitans"
/mol_type="mRNA"

```

```

/sub_species="morsitans"
/db_xref="taxon:37546"
/clone="Tse24g11_q1c"
/tissue_type="adult infected gut"
/clone_lib="Glossina morsitans morsitans adult infected
gut"
/note="Country: Zimbabwe; EST from adult gut infected with
T.brucei"

```

## ORIGIN

```

Query Match      84.8%; Score 17.8; DB 5; Length 443;
Best Local Similarity 90.5%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 TGCATGTATGGGTATCTTC 21
        ||||| ||||| ||||| |||||
DB      66 TCCATGTATGGGTATCGTCC 86

```

## RESULT 3

```

A1163927/c
LOCUS      A051P45U Hybrid aspen plasmid library Populus tremula x Populus
DEFINITION tremulooides cDNA 5', mRNA sequence.

```

```

ACCESSION  A1163927
VERSION     A1163927.1 GI:3955212
KEYWORDS   EST.
SOURCE     Populus tremula x Populus tremulooides

```

## ORGANISM

```

Populus tremula x Populus tremulooides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
Rosids; euroside I; Malpighiales; Salicaceae; Populus.

```

## REFERENCE

```

1 (bases 1 to 456)
Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A.,
Holmberg, A., Amini, B., Bhallerao, R., Larsson, M., Villarroel, R., Van
Montagu, M., Sandberg, G., Olsson, O., Teeri, T.T., Boerjan, W.,
Gustafsson, P., Uhlen, M., Sundberg, B. and Lundberg, J.

```

```

Gene discovery in the wood-forming tissues of poplar: Analysis of
5,692 expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
9789088
Contact: Sterky F

```

## JOURNAL

```

PUBMED     Department of Biotechnology
            Royal Institute of Technology (KTH)
            Teknikringen 34, S-100 44 STOCKHOLM, Sweden
            Tel: +46 8 790 8287
            Fax: +46 8 24 54 52
            Email: fredrik@biochem.kth.se

```

## COMMENT

```

PCR Primers
FORWARD: AAAGGGGATGTGCTGCAGGCG
BACKWARD: GCTCCGCTCGTATGTGTGTG
Seq primer: CGTTGTAACGACGCCAG
High quality sequence stop: 456.

```

## FEATURES

```

source      Location/Qualifiers
1..456
/organism="Populus tremula x Populus tremulooides"
/mol_type="mRNA"
/db_xref="taxon:47664"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/clone_lib="Hybrid aspen plasmid library"

```

```

/note="Vector: pBluescript SK; Site 1: SalI; Site 2: NotI;
Cambial region tissues, including developing xylem, the
meristematic cambial zone and the developing and mature
phloem, was harvested from 1.5 m actively growing trees.
cDNA was prepared and cloned into lambda gt22a. DNA was
isolated and subcloned into pBluescript SK using SalI and
NotI restriction enzymes."

```

## ORIGIN

```

Query Match      84.8%; Score 17.8; DB 1; Length 456;
Best Local Similarity 90.5%; Pred. No. 5.8e+02;

```

```

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21
    ||||| ||||| |||||
Db 419 TGCATGTAGGGTTACCTCC 399

RESULT 4
BG660206 483 bp mRNA linear EST 11-MAY-2001
LOCUS TgESTz297h11.y1 TgVEG118 Tachyzoite cDNA Library Toxoplasma gondii
DEFINITION cDNA clone TgESTz297h11.y1 5' similar to TR:Q9VHS8 Q9VHS8 CG7483
PROTEIN ; mRNA sequence.
ACCESSION BG660206
VERSION BG660206.1 GI:13803341
KEYWORDS EST.
SOURCE Toxoplasma gondii
ORGANISM Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
1 (bases 1 to 483)
Tang,K., Cole,R., Fogarty,S., Sibley,L.D., Ajioke,J.A., White,M.,
Clifton,S., Pape,D., Martin,J., Wyllie,T., Dante,M., Marra,M.,
Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M.,
Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I.,
Kennedy,S., Maguire,L., Waterston,R. and Wilson,R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.
Toxoplasma EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
Contact David Sibley (toxest@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco
High quality sequence stop: 369.
Location/Qualifiers
1. 483
/organism="Toxoplasma gondii"
/mol_type="mRNA"
/strain="VEG"
/db_xref="taxon:5811"
/clone="TgESTz297h11.y1"
/dev_stage="Tachyzoite"
/lab_host="DH10B"
/clone_lib="TgVEG118 Tachyzoite cDNA Library"
/note="Vector: pBluescript SK; Site1: EcoRI; Site2:
XhoI; This library was constructed by Keliang Tang, Robert
Cole, and L. David Sibley at Washington University. CDNA5
were synthesized from poly(A)+ RNA by o1idd(T) priming,
size-selected and directionally cloned into the Uni-ZAP XR
lambda vector (Stratagene). The primary library was mass
excised as phagemids and rescued in SOLR cells. The
plasmid library was recovered from the SOLR cells and
transformed in mass into DH10B (GeneHog, Research
Genetics, Inc.) for sequencing. WARNING: This library may
contain a small percentage contaminants from human
fibroblast cells."

ORIGIN
Query Match 84.8%; Score 17.8; DB 2; Length 483;
Best Local Similarity 90.5%; Pred. No. 5.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21
    ||||| ||||| |||||
Db 387 TGCATGTCTGGGTTCTTCC 407

RESULT 5

```

```

BH327941/c
LOCUS BH327941
DEFINITION CH230-117B21.TJ CHORI-230 Segment 1 Rattus norvegicus genomic clone
ACCESSION BH327941
VERSION BH327941.1 GI:17258655
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Muridae; Muridae; Muridae; Rattus.
1 (bases 1 to 560)
Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 EcoRI segment
Unpublished (1999)
Other GSSs: CH230-117B21.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 117 row: B column: 21
Seq primer: SP6
Class: BAC ends.
Location/Qualifiers
1. 560
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SsNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-117B21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 1"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN
Query Match 84.8%; Score 17.8; DB 9; Length 560;
Best Local Similarity 90.5%; Pred. No. 5.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21
    ||||| ||||| |||||
Db 546 TGTATGCTGGGTATCTTCC 526

RESULT 6
DE064462/c
LOCUS DE064462
DEFINITION Oryzias latipes DNA, clone: olal-111B07.R, genomic survey sequence.
ACCESSION DE064462
VERSION DE064462.1 GI:62576006
KEYWORDS GSS.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1

```

```

AUTHORS      Fujiyama, A., Toyoda, A., Kuroki, Y. and Sakaki, Y.
TITLE        BAC end sequences of Olal Oryzias latipes Library
JOURNAL      Published Only in Database (2005)
REFERENCE    2 (bases 1 to 619)
AUTHORS      Fujiyama, A.
TITLE        Direct Submission
JOURNAL      Submitted (12-APR-2005) Asao Fujiyama, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
              (E-mail: afujiyam@gsc.riken.jp, URL: http://stt.gsc.riken.jp/,
              Tel: 81-3-4212-2558, Fax: 81-3-3556-1916)
              This work was done in collaboration with Takeda, H. (1), Naruse, K.
              (2)
              and Narita, T. (3)
              (1) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: htakeda.s.u-tokyo.ac.jp
              (2) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: naruse.s.u-tokyo.ac.jp
              (3) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: tanarita.s.u-tokyo.ac.jp
              PRIMERS
              Sequencing : Forward
LIBRARY      Vector      : pKS145
              R.Site 1   : SacI
              L.Site 2   : SacI
              Location/Qualifiers
FEATURES     source
              1..619
              /organism="Oryzias latipes"
              /mol_type="genomic DNA"
              /db_xref="taxon:8090"
              /clone="olal-164P16.F"
              /sex="male"
              /cell_type="whole body"
              /clone_lib="BAC end sequences of Olal Oryzias latipes
              library"
ORIGIN
Query Match      84.8%; Score 17.8; DB 11; Length 619;
Best Local Similarity 90.5%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1   TGCATGTCATGGTTCCTCC 21
          |||||
Db      518 TGCATGTCATGGTTCCTCC 498

RESULT 7
LOCUS      DE067970
DEFINITION Oryzias latipes DNA, clone: olal-164P16.F, Genomic survey sequence.
ACCESSION  DE067970
VERSION     DE067970.1 GI:62579514
KEYWORDS    GSS.
SOURCE      Oryzias latipes (Japanese medaka)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
              Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
              Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
              1
REFERENCE   1

```

---

```

AUTHORS      Fujiyama, A., Toyoda, A., Kuroki, Y. and Sakaki, Y.
TITLE        BAC end sequences of Olal Oryzias latipes Library
JOURNAL      Published Only in Database (2005)
REFERENCE    2 (bases 1 to 637)
AUTHORS      Fujiyama, A.
TITLE        Direct Submission
JOURNAL      Submitted (12-APR-2005) Asao Fujiyama, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan
              (E-mail: afujiyam@gsc.riken.jp, URL: http://stt.gsc.riken.jp/,
              Tel: 81-3-4212-2558, Fax: 81-3-3556-1916)
              This work was done in collaboration with Takeda, H. (1), Naruse, K.
              (2)
              and Narita, T. (3)
              (1) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: htakeda.s.u-tokyo.ac.jp
              (2) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: naruse.s.u-tokyo.ac.jp
              (3) Department of Biological Science,
              University of Tokyo
              Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN
              Phone: +81-3-5841-4431
              Fax: +81-3-5841-4993
              E-mail: tanarita.s.u-tokyo.ac.jp
              PRIMERS
              Sequencing : Forward
LIBRARY      Vector      : pKS145
              R.Site 1   : SacI
              L.Site 2   : SacI
              Location/Qualifiers
FEATURES     source
              1..637
              /organism="Oryzias latipes"
              /mol_type="genomic DNA"
              /db_xref="taxon:8090"
              /clone="olal-164P16.F"
              /sex="male"
              /cell_type="whole body"
              /clone_lib="BAC end sequences of Olal Oryzias latipes
              library"
ORIGIN
Query Match      84.8%; Score 17.8; DB 11; Length 637;
Best Local Similarity 90.5%; Pred. No. 6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1   TGCATGTCATGGTTCCTCC 21
          |||||
Db      518 TGCATGTCATGGTTCCTCC 498

RESULT 8
LOCUS      BM645185/c
DEFINITION 1700687315834 A.Gam.ad.cdNAL Anopheles gambiae CDNA clone
ACCESSION  BM645185
VERSION     BM645185
KEYWORDS    EST.
SOURCE      Anopheles gambiae (African malaria mosquito)
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
              Culicidae; Anophelinae; Anopheles.
              1 (bases 1 to 714)

```



**AUTHORS**  
Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,  
Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.

**TITLE**  
Celera Anopheles gambiae EST project

**JOURNAL**  
Unpublished (2002)

**COMMENT**  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 24045334580  
Email: HoltR@celera.com  
Plate: N00100417U row: D column: 11  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .714  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced eusc. to Permethrin - std. chromosome)"  
/db\_xref="taxon:7165"  
/clones="19600449664157"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/clone\_lib="A.Gam.ad.cdna1"  
/note="vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

**ORIGIN**  
Query Match 84.8%; Score 17.8; DB 3; Length 714;  
Best Local Similarity 90.5%; Pred. No. 6.1e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**Qy** 1 TGCATGTATGGTGTATCTTCC 21  
|||||

**Db** 196 TGCATGTATGGTGTCTTCC 176  
|||||

**RESULT 9**  
DN793467 749 bp mRNA linear EST 15-APR-2005

**LOCUS**  
Pm\_Adms\_13F01\_T3 Helene and Hannah Pecten maximus cDNA clone

**DEFINITION**  
DN793467  
DN793467.1 GI:62630927

**ACCESSION**  
EST.

**KEYWORDS**  
Pecten maximus

**SOURCE**  
Pecten maximus

**ORGANISM**  
Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoidea; Pectinoidea; Pectinidae; Pecten.

**REFERENCE**  
1 (bases 1 to 749)  
Pagard, H.H., Andersen, O., Chubb, H.M.A. and Johnston, I.A.  
ESTs from adductor muscle of the Great Scallop (Pecten maximus) Unpublished (2005)

**AUTHORS**  
Contact: Ian A. Johnston  
Gatty Marine Laboratory, School of Biology  
University of St Andrews  
East Sands, St Andrews, Fife, KY16 8LB, Scotland, UK  
Tel: 44 (0)1334 463440  
Fax: 44 (0)1334 463008  
Email: iaj@st-and.ac.uk

**JOURNAL**  
PCR Primers

**COMMENT**  
FORWARD: T3 short  
BACKWARD: T7 short  
Plate: 13 row: F column: 01  
Seq primer: T3  
High quality sequence stop: 535.  
Location/Qualifiers  
1. .749  
/organism="Pecten maximus"  
/mol\_type="mRNA"

**FEATURES**  
source

**AUTHORS**  
Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,  
Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.

**TITLE**  
Celera Anopheles gambiae EST project

**JOURNAL**  
Unpublished (2002)

**COMMENT**  
Contact: Holt R.A.  
Celera Genomics  
45 W. Gude Dr., Rockville, MD 20850, USA  
Tel: 2404533151  
Fax: 24045334580  
Email: HoltR@celera.com  
Plate: N00100417U row: D column: 11  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .714  
/organism="Anopheles gambiae"  
/mol\_type="mRNA"  
/strain="RSP-ST (Reduced eusc. to Permethrin - std. chromosome)"  
/db\_xref="taxon:7165"  
/clones="19600449664157"  
/dev\_stage="Adult"  
/lab\_host="DH10B"  
/clone\_lib="A.Gam.ad.cdna1"  
/note="vector: pSport1; Site 1: SalI; Site 2: NotI; Whole adult mosquitoes (mixed sex) frozen on liquid nitrogen. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)."

**ORIGIN**  
Query Match 84.8%; Score 17.8; DB 8; Length 749;  
Best Local Similarity 90.5%; Pred. No. 6.2e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**Qy** 1 TGCATGTATGGTGTATCTTCC 21  
|||||

**Db** 572 TGCATGTATGGTGTATCTTCC 592  
|||||

**RESULT 10**  
DU058511 847 bp DNA linear GSS 12-AUG-2005

**LOCUS**  
90280 Tomato HindIII BAC Library Lycopersicon esculentum genomic clone LE\_HBA0150A21 3, genomic survey sequence.

**DEFINITION**  
DU058511  
DU058511.1 GI:72492098

**ACCESSION**  
GSS.

**VERSION**  
Lycopersicon esculentum (Solanum lycopersicum)

**KEYWORDS**  
Lycopersicon esculentum

**SOURCE**  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

**ORGANISM**  
1 (bases 1 to 847)  
Mueller, L.A., Buele, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J., Van Eck, J. and Stack, S.  
BAC end sequencing from three Solanum lycopersicon libraries Unpublished (2005)

**AUTHORS**  
Contact: Lukas Mueller  
Other GSSs: 90281  
Tanksley Lab, Dept. of Plant Breeding  
Cornell University  
251 Emerson Hall, Ithaca, NY 14853, USA  
Tel: 607-255-6557  
Fax: 607-255-6683  
Email: sgn-feedback@sgn.cornell.edu  
Plate: 150 row: A column: 21  
Seq primer: SP6  
Class: BAC ends  
High quality sequence start: 47  
High quality sequence stop: 434.  
Location/Qualifiers  
1. .847  
/organism="Lycopersicon esculentum"  
/mol\_type="genomic DNA"  
/cultivar="Heinz 1706"  
/db\_xref="taxon:4081"  
/clones="LE\_HBA0150A21"  
/lab\_host="E. coli"  
/clone\_lib="Tomato HindIII BAC Library"  
/note="vector: pBelOBAC11; Site\_1: HindIII"

**TITLE**  
JOURNAL

**COMMENT**

**FEATURES**  
source

**ORIGIN**  
Query Match 84.8%; Score 17.8; DB 10; Length 847;  
Best Local Similarity 90.5%; Pred. No. 6.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

**Qy** 1 TGCATGTATGGTGTATCTTCC 21  
|||||

**Db** 435 TGCATGTATGGTGTCTTCC 455  
|||||

**RESULT 11**  
CZ954709 933 bp DNA linear GSS 11-AUG-2005

**LOCUS**  
268426 Tomato EcoRI BAC Library Lycopersicon esculentum genomic clone LE\_HBA0150A21 3, genomic survey sequence.

**DEFINITION**  
CZ954709  
CZ954709.1 GI:100000000

**ACCESSION**  
GSS.

**VERSION**  
Lycopersicon esculentum

**KEYWORDS**  
Lycopersicon esculentum

**SOURCE**  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

**ORGANISM**  
1 (bases 1 to 933)  
Mueller, L.A., Buele, R.M., Wang, Y., Tanksley, S.D., Giovannoni, J.J., Van Eck, J. and Stack, S.  
BAC end sequencing from three Solanum lycopersicon libraries Unpublished (2005)

**AUTHORS**  
Contact: Lukas Mueller  
Other GSSs: 90281  
Tanksley Lab, Dept. of Plant Breeding  
Cornell University  
251 Emerson Hall, Ithaca, NY 14853, USA  
Tel: 607-255-6557  
Fax: 607-255-6683  
Email: sgn-feedback@sgn.cornell.edu  
Plate: 150 row: A column: 21  
Seq primer: SP6  
Class: BAC ends  
High quality sequence start: 47  
High quality sequence stop: 434.  
Location/Qualifiers  
1. .933  
/organism="Lycopersicon esculentum"  
/mol\_type="genomic DNA"  
/cultivar="Heinz 1706"  
/db\_xref="taxon:4081"  
/clones="LE\_HBA0150A21"  
/lab\_host="E. coli"  
/clone\_lib="Tomato HindIII BAC Library"  
/note="vector: pBelOBAC11; Site\_1: HindIII"

**TITLE**  
JOURNAL

**COMMENT**

**FEATURES**  
source





```

QY      1  TGCATGTATGGGTTATCTT 19
      ||||| ||||| ||||| |||||
Db      43  TGCATATATGGGTTATCTT 25

RESULT 16
AW489071/c
LOCUS   397 bp      mRNA      linear      EST 24-FEB-2000
DEFINITION  UI-M-BH3-asd-h-03-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
ACCESSION  UI-M-BH3-asd-h-03-0-UI 3', mRNA sequence.
VERSION    AW489071
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus

REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muroidea; Muridae; Murinae; Mus.
TITLE      1 (bases 1 to 397)
JOURNAL    Bonaldo,M.F., Lennon,G. and Soares,M.B.
PUBMED     8889548
COMMENT    Normalization and subtraction: two approaches to facilitate gene
           discovery
           Genome Res. 6 (9), 791-806 (1996)
           Contact: Chin, H
           National Institute of Mental Health
           6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
           20892-9643, USA
           Tel: 301 443 1706
           Fax: 301 443 9890
           Email: mEST@mail.nih.gov
           Oligo-dT track not found, Not I site shown in beginning of sequence
           is likely internal to the message. cDNA Library Preparation: M.B.
           Soares Lab Clone Distribution: Researchers may obtain BMAP cDNA
           clones from RESEARCH GENETICS. It should be noted that Bento Soares
           is generating a small number of additional specialized
           non-redundant arrays of BMAP cDNAs whose availability will be
           considered under appropriate and limited collaborative arrangements
           Seq primer: M13 Forward
           POLYA=No.

FEATURES
source
1. .397
   /organism="Mus musculus"
   /mol_type="mRNA"
   /strain="C57BL/6J"
   /db_xref="taxon:10090"
   /clone="UI-M-BH3-asd-h-03-0-UI"
   /dev_stage="27-32 days"
   /lab_host="DH10B (Life Technologies)"
   /clone_lib="NIH_BMAP_M_S4"
   /notes="Vector: pF773D-Pac (Pharmacia) with a modified
   polylinker; Site 1: Not I; Site 2: Eco RI; The
   NIH_BMAP_M_S4 library is a subtracted library of a series,
   ultimately derived from a mixture of individually tagged
   normalized libraries from ten regions of the mouse brain
   (cerebellum, brain stems, olfactory bulbs, hypothalamus,
   cortex, amygdala, basal ganglia, pineal gland, striatum,
   hippocampus) after a series of subtractions to reduce the
   representation of cDNAs from which ESTs had already been
   generated. The following serially subtracted libraries
   were generated in this process: NIH_BMAP_M_S4,
   NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1,
   NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library
   (NIH_BMAP_M_S4) was constructed as follows: PCR amplified
   cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and
   NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived
   was used as a driver in a hybridization with a pool of
   the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1
   libraries in the form of single-stranded circles. The
   remaining single-stranded circles (subtracted library)
   was purified by hydroxyapatite column chromatography,
   converted to double-stranded circles and electroporated
   into DH10B bacteria (LifeTechnologies) to generate the

```

---

```

ORIGIN
Query Match      82.9%; Score 17.4; DB 1; Length 397;
Best Local Similarity 94.7%; Pred. No. 9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  GCATGTATGGGTTATCTTC 20
      ||||| ||||| ||||| |||||
Db      91  GCATGTATGGGTTATCTTC 73

RESULT 17
AK219962
LOCUS    403 bp      mRNA      linear      HTC 23-NOV-2004
DEFINITION  Mus musculus cDNA, clone: Y2G0150E15, strand: minus,
           reference: ENSEMBL: Mouse-Transcript-ENST: ENSMUST00000061642, based
           on BLAT search.
ACCESSION  AK219962.1 GI: 56044353
VERSION    HTC; ASSETS.
KEYWORDS   Mus musculus (house mouse)
SOURCE     Mus musculus
ORGANISM   Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
           Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS    Watahiki, A., Waki, K., Hayatsu, N., Shiraki, T., Kondo, S.,
           Nakamura, M., Sasaki, D., Arakawa, T., Kawai, J., Harbers, M.,
           Hayashizaki, Y. and Carninci, P.
TITLE      Libraries enriched for alternatively spliced exons reveal splicing
           patterns in melanocytes and melanomas
JOURNAL    Nat. Methods 1, 233-239 (2004)
PUBMED     15404402
REFERENCE  2 (bases 1 to 403)
AUTHORS    Arakawa, T., Carninci, P., Fukuda, S., Harbers, M., Hayatsu, N.,
           Hori, F., Imotani, K., Kawai, J., Kondo, S., Murata, M., Nakamura, M.,
           Nomura, K., Ohno, M., Sasaki, D., Shiraki, T., Waki, K., Watahiki, A. and
           Hayashizaki, Y.
TITLE      Direct Submission
JOURNAL    Submitted (15-SEP-2004) Yoshihide Hayashizaki, The Institute of
           Physical and Chemical Research (RIKEN), Laboratory for Genome
           Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
           RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
           Kanagawa, 230-0045, Japan [E-mail: genome-res@gsc.riken.jp,
           URL: http://genome.gsc.riken.jp, Tel: 81-45-503-9222,
           Fax: 81-45-503-9216]
COMMENT    Alternative Splicing Libraries (ASLs) are prepared by: Preparing
           of single-stranded DNA using a RNA template from full length cDNA
           libraries, hybridizing of
           single-stranded DNAs, removing of remaining single-stranded DNA,
           bp-cutters, capturing of DNA hybrids with loop structures
           (alternative spliced exon), ligating of Y-shaped primers to
           isolated DNA hybrids with loop structures, PCR amplification of
           ligation products and their cloning into pFLCI vector. (Reference).
           Location/Qualifiers
           1..403
             /organism="Mus musculus"
             /mol_type="mRNA"
             /db_xref="taxon:10090"
             /clone="Y2G0150E15"
             /cell_lines="mixture of B16-F10Y and melan-c"
             /cell_type="mixture of melanoma cell and melanocyte cell"
             /clone_lib="Alternative Splicing Library L3"
             /note="strand: minus, reference: ENSEMBL: Mouse-Transcript-
             ENST: ENSMUST00000061642, based on BLAT search"

```

NIH\_BMAP\_M\_S4 library. This procedure has been previously described [Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996]  
TAG\_TISSUE=hypothalamus  
TAG\_LIB=NIH\_BMAP\_M\_S4  
TAG\_SEQ=CGGTA"

Query Match 82.9%; Score 17.4; DB 4; Length 403;  
 Best Local Similarity 94.7%; Pred. No. 9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 366 GCATGTATGGTTATGTTTC 384

RESULT 18  
 CB799737  
 LOCUS CB799737 416 bp mRNA linear EST 16-MAY-2003  
 DEFINITION AMGNNUC:NRW3-00059-H10-A white adipose tiss (10469) Rattus  
 norvegicus cDNA clone nrwa3-00059-h10 5', mRNA sequence.

ACCESSION CB799737  
 VERSION CB799737.1 GI:29911338

KEYWORDS EST.  
 SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Rattus.

REFERENCE 1 (bases 1 to 416)  
 AUTHORS Angen EST Program  
 TITLE Angen Rat EST Program  
 JOURNAL Unpublished (2003)  
 COMMENT Contact: Dan Fitzpatrick  
 Angen, Inc  
 One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA  
 Tel: 805 447-4881  
 Plate: 00059 row: h column: 10.

FEATURES  
 source  
 1..416  
 Location/Qualifiers

/organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="nrwa3-00059-h10"  
 /tissue\_type="adipose tiss"  
 /clone\_lib="white adipose tiss (10469)"  
 /note="vector: pSPORtl; Site\_1: SalI; Site\_2: NotI; white  
 adipose tiss adult female Wistar rats, >250 grams"

## ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 416;  
 Best Local Similarity 94.7%; Pred. No. 9.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 151 GCATGTATGGTTATGTTTC 169

RESULT 19  
 AW492409/c  
 LOCUS AW492409 429 bp mRNA linear EST 24-FEB-2000  
 DEFINITION UI-M-BH3-atu-c-05-0-UI s1 NIH BMAP\_M\_S4 Mus musculus cDNA clone  
 UI-M-BH3-atu-c-05-0-UI 3', mRNA sequence.

ACCESSION AW492409  
 VERSION AW492409.1 GI:7062690

KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 429)  
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)

COMMENT 8889548  
 Contact: Chin, H  
 National Institute of Mental Health

6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
 20892-9643, USA  
 Tel: 301 443 1706  
 Fax: 301 443 9890  
 Email: mEST@mail.nih.gov  
 Oligo-dt track not found, Not 1 site shown in beginning of sequence  
 is likely internal to the message. cDNA Library Preparation: M.B.  
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
 clones from RESEARCH GENETICS. It should be noted that Bento Soares  
 is generating a small number of additional specialized  
 non-redundant arrays of BMAP cDNAs whose availability will be  
 considered under appropriate and limited collaborative arrangements  
 Seq primer: M13 Forward  
 POLYA=No.

FEATURES  
 source

Location/Qualifiers

1..429  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UI-M-BH3-atu-c-05-0-UI"  
 /dev\_stage="27-32 days"  
 /lab\_host="DH10B (Life Technologies)"  
 /clone\_lib="NIH\_BMAP\_M\_S4"  
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
 polylinker; Site 1: Not I; Site 2: Eco RI; The  
 NIH BMAP\_M\_S4 library is a subtracted library of a series,  
 ultimately derived from a mixture of individually tagged  
 normalized libraries from ten regions of the mouse brain  
 (cerebellum, brain stems, olfactory bulbs, hypothalamus,  
 cortex, amygdala, basal ganglia, pineal gland, striatum,  
 hippocampus) after a series of subtractions to reduce the  
 representation of cDNAs from which ESTs had already been  
 generated. The following serially subtracted libraries  
 were generated in this process: NIH\_BMAP\_M\_S4,  
 NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, NIH\_BMAP\_M\_S3.1,  
 NIH\_BMAP\_M\_S2, NIH\_BMAP\_M\_S1. The subtracted library  
 (NIH\_BMAP\_M\_S4) was constructed as follows: PCR amplified  
 cDNA inserts from NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and  
 NIH\_BMAP\_M\_S3.1 clones from which 3' ESTs had been derived  
 was used as a driver in a hybridization with a pool of  
 the NIH\_BMAP\_M\_S3.3, NIH\_BMAP\_M\_S3.2, and NIH\_BMAP\_M\_S3.1  
 libraries in the form of single-stranded circles. The  
 remaining single-stranded circles (subtracted library)  
 was purified by hydroxyapatite column chromatography,  
 converted to double-stranded circles and electroporated  
 into DH10B bacteria (Life Technologies) to generate the  
 NIH\_BMAP\_M\_S4 library. This procedure has been previously  
 described (Bonaldo, Lennon and Soares, Genome Research  
 6:791-806, 1996)  
 TAG\_TISSUE=pineal-glands  
 TAG\_LIB=NIH\_BMAP\_M\_S4  
 TAG\_SEQ=CAGAC

## ORIGIN

Query Match 82.9%; Score 17.4; DB 1; Length 429;  
 Best Local Similarity 94.7%; Pred. No. 9.1e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 91 GCATGTATGGTTATGTTTC 73

RESULT 20  
 CB738293

LOCUS CB738293 459 bp mRNA linear EST 11-APR-2003  
 DEFINITION AMGNNUC:NRHY5-00423-B4-A W Rat hypothalamus (10471) Rattus  
 norvegicus cDNA clone nrhy5-00423-b4 5', mRNA sequence.

ACCESSION CB738293  
 VERSION CB738293.1 GI:29805558  
 KEYWORDS EST.  
 SOURCE Rattus norvegicus (Norway rat)

```
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.

REFERENCE
AUTHORS      Angen EST Program.
TITLE        Angen Rat EST Program
JOURNAL      Unpublished (2003)
COMMENT      Contact: Dan Fitzpatrick
Angen, Inc
One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
Tel: 805 447-4881
Plate: 00423 row: b column: 4.

FEATURES
source       Location/Qualifiers
1. .459
/mol_type="mRNA"
/db_xref="taxon:10116"
/clone="nrh5-00423-b4"
/clone_lib="W Rat hypothalamus (10471)"
/notes="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; W Rat
hypothalamus adult female Wistar rat avg. insert size 2.3
kb fraction 6 and 7"

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 459;
Best Local Similarity 94.7%; Pred. No. 9.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GCATGTATGGTTATCTTC 20
        |||||
Db      268 GCATGTATGGTTATGTTTC 286
        |||||

RESULT 21
LOCUS    CA895102
DEFINITION 481 bp mRNA linear EST 20-DEC-2002
Library (Long) Mus musculus cDNA clone NIA:B0189H07 IMAGE:30102426
5', mRNA sequence.
ACCESSION CA895102
VERSION    CA895102.1 GI:27346651
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 481)
AUTHORS   Piao,Y., Dudekula,D.B., Qian,Y., Martin,P.R., Aiba,K., Vescovi,A.L.
and Ko,M.S.H.
TITLE     Systematic Analyses of NIA Mouse Neural Stem Cell (Differentiated)
cDNA Library (Long)
JOURNAL   Unpublished (2002)
COMMENT   Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: B0189 row: H column: 07
Seq primer: -21M13 Reverse
High quality sequence stop: 481
POLYA=No.

FEATURES
source       Location/Qualifiers
1. .481
/mol_type="mRNA"
/strain="CD1"
/db_xref="niaEST:B0189H07-5N"
/db_xref="taxon:10090"
/clone="NIA:B0189H07 IMAGE:30102426"
/dev_stage="Adult"
/lab_host="DH10B"

/clone_lib="NIA Mouse Neural Stem Cell (Differentiated)
cDNA Library (Long)"
/notes="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Dr. Angelo L. Vescovi (Institute for Stem
Cell Research, Italy). Double-stranded cDNAs were
synthesized with an Oligo(dT) primer [Invitrogen:
5'-pGACTAGTTCGAGCGGCGCCCTTTTTC-3'] from
2.0 Microgram of total RNA, treated with T4 DNA
polymerase, and purified by ethanol-precipitation. The
cDNAs were ligated to lone-linker LL-Sal4, purified by
phenol/chloroform, and separated from free linkers by
Centricon 100. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer Sal4-S. The products were purified
by phenol/chloroform and Centricon 100. The cDNAs were
digested with SalI and NotI enzymes and cloned into
SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
coli host was transformed with the ligation mixture by the
standard chemical method. The average insert size is about
3.2 kb. The library was constructed by Yulan Piao."

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 481;
Best Local Similarity 94.7%; Pred. No. 9.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 GCATGTATGGTTATCTTC 20
        |||||
Db      277 GCATGTATGGTTATGTTTC 295
        |||||

RESULT 22
LOCUS    CA888060
DEFINITION 553 bp mRNA linear EST 20-DEC-2002
Library (Long) Mus musculus cDNA clone NIA:B0142D05 IMAGE:30097864
5', mRNA sequence.
ACCESSION CA888060
VERSION    CA888060.1 GI:27339609
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 553)
AUTHORS   Piao,Y., Dudekula,D.B., Qian,Y., Martin,P.R., Aiba,K., Vescovi,A.L.
and Ko,M.S.H.
TITLE     Systematic Analyses of NIA Mouse Neural Stem Cell (Differentiated)
cDNA Library (Long)
JOURNAL   Unpublished (2002)
COMMENT   Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: B0142 row: D column: 05
Seq primer: -21M13 Reverse
High quality sequence stop: 553
POLYA=No.

FEATURES
source       Location/Qualifiers
1. .553
/mol_type="mRNA"
/organism="Mus musculus"
/strain="CD1"
/db_xref="niaEST:B0142D05-5N"
/db_xref="taxon:10090"
/clone="NIA:B0142D05 IMAGE:30097864"
```

```

/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="NIA Mouse Neural Stem Cell (Differentiated)
cDNA Library (long)"
/notes="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://igeun-grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
obtained from Dr. Angelo L. Vescovi (Institute for Stem
Cell Research, Italy). Double-stranded cDNAs were
synthesized with an Oligo(dT) primer [Invitrogen:
5'-pGACGTAGTCTAGATCGGAGCGCGCCCTTTTCTTTT-3'] from
2.0 Microgram of total RNA, treated with T4 DNA
polymerase, and purified by ethanol-precipitation. The
cDNAs were ligated to lone-linker LL-Sal4, purified by
phenol/chloroform, and separated from free linkers by
Centricon 100. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer Sal4-S. The products were purified
by phenol/chloroform and Centricon 100. The cDNAs were
digested with SalI and NotI enzymes and cloned into
SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
coli host was transformed with the ligation mixture by the
standard chemical method. The average insert size is about
3.2 kb. The library was constructed by Yulan Piao."

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 553;
Best Local Similarity 94.7%; Pred. No. 9.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGTATGGTTATCTTC 20
      |||||
Db      277 GCATGTATGGTTATGTC 295

RESULT 23
CA637985      561 bp mRNA linear EST 23-NOV-2002
LOCUS      wreln.pk0005.h2 wreln Triticum aestivum cDNA clone wreln.pk0005.h2
DEFINITION      5' end, mRNA sequence.
ACCESSION      CA637985
VERSION      CA637985.1 GI:25216281
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.
REFERENCE      1 (bases 1 to 561)
AUTHORS      Tingey,S.V., Powell,W., Walters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE      Dupont Wheat cDNA Sequence
JOURNAL      Unpublished (2002)
COMMENT      Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.

FEATURES
source
Location/Qualifiers
1..561
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wreln.pk0005.h2"
/tissue_type="root"
/clone_lib="wreln"
/notes="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:

XhoI; Wheat (Triticum aestivum L.) root; normalized from
wre1 library"

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 561;
Best Local Similarity 94.7%; Pred. No. 9.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 CATGTATGGTTATCTTCC 21
      |||||
Db      301 CATGTATGGTTATCTACC 319

RESULT 24
BQ571868      581 bp mRNA linear EST 19-JUN-2002
LOCUS      UI-M-FCO-byc-b-20-0-UI.r1 NIH_BMAP_FCO Mus musculus cDNA clone
DEFINITION      IMAGE:5716147 5', mRNA sequence.
ACCESSION      BQ571868
VERSION      BQ571868.1 GI:21475185
KEYWORDS      EST.
SOURCE      Mus musculus (house mouse)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 581)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.

FEATURES
source
Location/Qualifiers
1..581
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="C57BL/6"
/clone="IMAGE:5716147"
/tissue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH_BMAP_FCO"
/notes="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA
tail, is TGAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 5; Length 581;
Best Local Similarity 94.7%; Pred. No. 9.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```



```

QY      2 GCATGATGGGTATCTTC 20
DB      226 GCATGATGGGTATCTTC 244

RESULT 25
BQ571429
LOCUS   589 bp mRNA linear EST 19-JUN-2002
DEFINITION
        UI-M-FCO-byb-f-23-0-UI.r1 NIH_BMAP_FCO Mus musculus cDNA clone
        IMAGE:5715862 5', mRNA sequence.

ACCESSION
BQ571429
VERSION
BQ571429.1 GI:21474746
KEYWORDS
Esf.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 589)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-rc@mail.nih.gov
Tissue procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source
       1..589
         /organism="Mus musculus"
         /mol_type="mRNA"
         /strain="C57BL/6"
         /db_xref="taxon:10090"
         /clone="IMAGE:5715862"
         /tissue_type="whole brain"
         /dev_stage="embryo 12.5 dpc"
         /lab_host="DH10B (T1 phage resistant)"
         /clone_lib="NIH_BMAP_FCO"
         /notes="Organ: brain; Vector: pYX-Asc; Site 1: Ecor I;
         Site 2: Not I; The library was constructed according to
         Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
         1996. Denatured mRNA was size fractionated on a 1% agarose
         gel. First strand cDNA synthesis was primed with an
         oligo-dT primer containing a Not I site. Double stranded
         cDNA was size selected according to mRNA size fraction,
         ligated with Ecor I adaptor, digested with Not I, and then
         cloned directionally into pYX-Asc vector. The library tag
         sequence located between the Not I site and the polyA
         tail, is TGAGAGAGCC. This library was created for the
         University of Iowa Mouse Brain Molecular Anatomy Project
         (BMAP). 'Gene Discovery in the Developing Mouse Nervous
         System', supported by National Institutes of Mental Health
         (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 5; Length 589;
Best Local Similarity 94.7%; Pred. No. 9.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGATGGGTATCTTC 20
DB      331 GCATGATGGGTATCTTC 349

RESULT 26

```

```

CE495411/c
LOCUS   592 bp DNA linear GSS 28-SEP-2003
DEFINITION
        tigr-gss-dog-17000310260727 Dog Library Canis familiaris genomic,
        genomic survey sequence.

ACCESSION
CE495411
VERSION
CE495411.1 GI:36812192
KEYWORDS
GSS.
SOURCE
Canis familiaris (dog)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.

REFERENCE
1 (bases 1 to 592)
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Frazer,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirkness@tigr.org
Class: shotgun.
Location/Qualifiers
     source
       1..592
         /organism="Canis familiaris"
         /mol_type="genomic DNA"
         /strain="Standard Poodle"
         /db_xref="taxon:9615"
         /clone_lib="Dog Library"
         /note="Site 1: BstXI; Libraries were prepared from
         peripheral blood"

ORIGIN
Query Match      82.9%; Score 17.4; DB 10; Length 592;
Best Local Similarity 94.7%; Pred. No. 9.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTT 19
DB      94 TGCATGTATGGGTATGTT 76

RESULT 27
BZ300676/c
LOCUS   594 bp DNA linear GSS 31-OCT-2002
DEFINITION
        BZ300676
        X00856.p1 Kluyveromyces delphensis Random Genomic Library
        Kluyveromyces delphensis genomic clone X00856, genomic survey
        sequence.

ACCESSION
BZ300676
VERSION
BZ300676.1 GI:24445706
KEYWORDS
GSS.
SOURCE
Kluyveromyces delphensis
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

REFERENCE
1 (bases 1 to 594)
Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
Evidence from comparative genomics for a complete sexual cycle in
the 'asexual' pathogenic yeast Candida glabrata
Genome Biol. 4 (2), R10 (2003)
12620120
Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie

```



```

FEATURES
  source
    Class: plasmid ends.
    Location/Qualifiers
      1..594
        /organism="Kluyveromyces delphensis"
        /mol_type="genomic DNA"
        /strain="CBS 2170"
        /db_xref="taxon:51657"
        /clone="KD0856"
        /clone_lib="Kluyveromyces delphensis Random Genomic Library"

ORIGIN
  Query Match      82.9%; Score 17.4; DB 9; Length 594;
  Best Local Similarity 94.7%; Pred. No. 9.5e+02;
  Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTT 19
    ||||| ||||| |||||
Db 411 TGCATGATGGGTATCTT 393

RESULT 28
CE104077      596 bp DNA linear GSS 24-SEP-2003
LOCUS tigr-gss-dog-17000371142608 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE104077
VERSION CE104077.1 GI:35170962
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
          Canis.
REFERENCE 1 (bases 1 to 596)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
          The Institute for Genomic Research
          Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
          Rockville, MD 20850, USA
          Tel: 301-838-0200
          Fax: 301-838-0208
          Email: ekirknes@tigr.org
          Class: shotgun.
          Location/Qualifiers
            1..596
              /organism="Canis familiaris"
              /mol_type="genomic DNA"
              /strain="Standard Poodle"
              /db_xref="taxon:9615"
              /clone_lib="Dog Library"
              /note="Site 1: BstXI; Libraries were prepared from
              peripheral blood"

FEATURES
  source
    Query Match      82.9%; Score 17.4; DB 9; Length 596;
    Best Local Similarity 94.7%; Pred. No. 9.5e+02;
    Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGATGGGTATCTT 19
    ||||| ||||| |||||
Db 191 TGCATGATGGGTATCTT 209

RESULT 29
BE336403      600 bp mRNA linear EST 14-JUL-2000
LOCUS tigr-gss-dog-17000371142608 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE133541
VERSION CE133541.1 GI:35235919
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
          Canis.
REFERENCE 1 (bases 1 to 643)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
          The Institute for Genomic Research

```

```

mRNA sequence.
BE336403      643 bp DNA linear GSS 25-SEP-2003
LOCUS tigr-gss-dog-17000371142608 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE133541
VERSION CE133541.1 GI:35235919
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
          Canis.
REFERENCE 1 (bases 1 to 643)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
          The Institute for Genomic Research

```

```

mRNA sequence.
BE336403      643 bp DNA linear GSS 25-SEP-2003
LOCUS tigr-gss-dog-17000371142608 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE133541
VERSION CE133541.1 GI:35235919
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
          Canis.
REFERENCE 1 (bases 1 to 643)
AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
          Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
          Venter,J.C.
TITLE The dog genome: survey sequencing and comparative analysis
JOURNAL Science 301 (5641), 1898-1903 (2003)
PUBMED 14512627
COMMENT Contact: Kirkness EF
          The Institute for Genomic Research

```

Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

# FEATURES

## source

Location/Qualifiers  
1. .643  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from  
peripheral blood"

## ORIGIN

Query Match 82.9%; Score 17.4; DB 9; Length 643;  
Best Local Similarity 94.7%; Pred. No. 9.6e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTT 19  
|||||  
DB 570 TGCATGTATGGGTATGTT 588

## RESULT 31

### BE533917

#### LOCUS

DEFINITION BE533917 685 bp mRNA linear EST 09-AUG-2000  
161234117F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:3597604 5',  
mRNA sequence.

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
1 (bases 1 to 685)

## REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### COMMENT

NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM8777 row: b column: 05  
High quality sequence stop: 586.

## FEATURES

### source

Location/Qualifiers  
1. .685  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3597604"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Mam6"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

## ORIGIN

Query Match 82.9%; Score 17.4; DB 2; Length 685;

Best Local Similarity 94.7%; Pred. No. 9.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 20  
|||||  
DB 17 GCATGTATGGGTATGTT 35

## RESULT 32

### CA318453

#### LOCUS

DEFINITION CA318453 691 bp mRNA linear EST 09-JUL-2003  
UI-M-FW0-cbs-i-10-0-UI.r1 NIH BMAP\_FW0 Mus musculus cDNA clone  
IMAGE:6813875 5', mRNA sequence.

#### ACCESSION

#### VERSION

#### KEYWORDS

#### SOURCE

#### ORGANISM

Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
1 (bases 1 to 691)

## REFERENCE

### AUTHORS

### TITLE

### JOURNAL

### COMMENT

NIH-MGC http://mgi.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

## Seq primer: pYX-5.

### Location/Qualifiers

#### source

1. .891  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6813875"  
/tissue\_type="whole brain"  
/dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP FW0"  
/note="Organ: Brain; Vector: pYX-Aac; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction. ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Aac vector. The library tag  
sequence located between the Not I site and the polyA tail  
is AGCAGACAG. This library was created for the University  
of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
program coordinator."

Query Match 82.9%; Score 17.4; DB 6; Length 691;  
Best Local Similarity 94.7%; Pred. No. 9.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## ORIGIN

QY 2 GCATGTATGGGTATCTTC 20  
|||||  
DB 56 GCATGTATGGGTATGTT 74

```

RESULT 33
CF749756
LOCUS
DEFINITION
  UI-M-HJO-cmn-o-24-0-UI.r1 NIH BMAP_HJO Mus musculus cDNA clone
  IMAGE:30628103 5', mRNA sequence.
ACCESSION
CF749756
VERSION
CF749756.1 GI:37646101
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 697)
NTH-MGC http://mgc.nci.nih.gov/.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgapbs-x@mail.nih.gov
Tissue Procurement: Dr. James Lin University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
http://genome.uiowa.edu/distribution/mousefl.html
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
FEATURES
    source
        Location/Qualifiers
            1..697
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6"
                /db_xref="taxon:10090"
                /clone="IMAGE:30628103"
                /tissue_type="Upper Head"
                /dev_stage="9.5 and 10.5 dpc"
                /lab_host="DH10B (TI phage resistant)"
                /clone_lib="NIH BMAP_HJO"
                /note="Organ: Head; Vector: pYX-Asc; Site 1: EcoR I;
                Site 2: Not I; The library was constructed according
                to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                1996. Denatured RNA was size fractionated on a 1% agarose
                gel. First strand cDNA synthesis was primed with oligo-dT
                primer containing a Not I site. Double strand cDNA was
                size selected according to mRNA size fraction, ligated
                with EcoR I adaptor, digested with NotI and then cloned
                directionally into pYX-Asc vector. The library tag
                sequence located between the Not I site and the polyA tail
                is CGAACTGAAT. This library was created for the University
                of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
                Developing Mouse Nervous System', supported by National
                Institute of Mental Health (NIMH)."
```

```

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 697;
Best Local Similarity 94.7%; Pred. No. 9.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy  2 GCATGTATGGGTATCTTC 20
    |||||
Db  57 GCATGTATGGGTATGTTC 75

```

```

RESULT 34
CAS111105
LOCUS
DEFINITION
  UI-R-FJ0-cpv-e-20-0-UI.r1 UI-R-FJ0 Rattus norvegicus cDNA clone
  IMAGE:6838739 5', mRNA sequence.
ACCESSION
CAS111105
VERSION
CAS111105.1 GI:25002059
KEYWORDS
EST.
```

## SOURCE ORGANISM

Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Rattus.

## REFERENCE AUTHORS TITLE

1 (bases 1 to 702)  
Bonaldo,M.F., Lennon,G. and Soares,M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery

## JOURNAL PUBMED COMMENT

Genome Res. 6 (9), 791-806 (1996)

8889548

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565

Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
Seq primer: M13 REVERSE.

## FEATURES source

1..702  
Location/Qualifiers

/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-FJ0-cpv-e-20-0-UI"  
/tissue\_type="embryo"  
/dev\_stage="embryo"  
/lab\_host="DH10B (Life Technologies) (TI phage resistant)"  
/clone\_lib="UI-R-FJ0"  
/notes="Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I;  
UI-R-FJ0 is a cDNA library containing the following  
tissue(s): rat embryo. The library was constructed  
according to Bonaldo, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an EcoR I adaptor, digested  
with Not I, and cloned directionally into pYX-Asc  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is CATCTTACT. This library  
was created for the University of Iowa Program for Rat  
Gene Discovery and Mapping (Val Sheffield, Bento Soares  
and Tom Casavant)"

## ORIGIN

```

Query Match      82.9%; Score 17.4; DB 6; Length 702;
Best Local Similarity 94.7%; Pred. No. 9.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy  2 GCATGTATGGGTATCTTC 20
    |||||
Db  129 GCATGTATGGGTATGTTC 147

```

## RESULT 35

CBS19329

LOCUS

DEFINITION

UI-M-GH0-cef-e-10-0-UI.r1 NIH BMAP\_GH0 Mus musculus cDNA clone

IMAGE:6838739 5', mRNA sequence.

ACCESSION

CB519329

VERSION

CB519329.1 GI:29352684

KEYWORDS

Mus musculus (house mouse)

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 707)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Jim Lin, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

FEATURES  
source

Location/Qualifiers

1..707  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6838739"  
/tissue\_type="Whole brain"  
/dev\_stage="1, 5, and 15 days newborn"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP GH0"  
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: Ecor I; Site 2: Not I; The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with Ecor I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CGAAGCTGAAT. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., Program coordinator."

ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 707;  
Best Local Similarity 94.7%; Pred. No. 9.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTATCTTC 20

Db 170 GCATGTATGGTATCTTC 188

RESULT 36  
CV805733/c

LOCUS

DEFINITION CV805733 714 bp mRNA linear EST 21-JAN-2005

IMAGE:7465640 5', mRNA sequence.

ACCESSION CV805733

VERSION CV805733.1 GI:55748699

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM Xenopus tropicalis

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 714)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH  
Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)

Tissue Procurement: Bruce Blumberg

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM15760 row: b column: 06

High quality sequence start: 6

High quality sequence stop: 282.

FEATURES

source

Location/Qualifiers

1..714

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="N6 (Nigerian 6th generation inbred)"

/db\_xref="taxon:8364"

/clone="IMAGE:7465640"

/tissue\_type="Kidney"

/dev\_stage="Adult"

/lab\_host="ElectroMAX DH10B T1 Phage Resistant cells"

/clone\_lib="NIH XGC tropKidi"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The

library was prepared from 5 ug of poly A+ RNA by oligo-dT

priming

(5'-GAGAGAGAGAGAGAGACTAGTCTCGAGTTTTTTTTTTTTTT-3')

and Stratascript reverse transcriptase. After ligation of

EcoRI adapters (5'-AATTCGACAGG-3') followed by kinasing

adapters and by XhoI digestion, the cDNA was size selected

by chromatography on Sepharose CL-2B columns and fractions

containing cDNAs larger than 1000 bp were ligated into

EcoRI/XhoI-digested pCS107. Reference for library

construction: Current Genomics 4, 635-644. Library

constructed by Michelle Tabb and Bruce Blumberg (Dept of

Developmental and Cell Biology, University of California,

Irvine).

ORIGIN

Query Match 82.9%; Score 17.4; DB 8; Length 714;  
Best Local Similarity 94.7%; Pred. No. 9.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTATCTT 19

Db 359 TGCATGTATGTATCTT 341

RESULT 37

CK364526

LOCUS

DEFINITION CK364526 721 bp mRNA linear EST 23-DEC-2003

IMAGE:7096948 5', mRNA sequence.

ACCESSION CK364526

VERSION CK364526.1 GI:40330461

KEYWORDS EST.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 721)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM14942 row: h column: 02  
 High quality sequence stop: 722.  
 Location/Qualifiers

#### FEATURES

1..721  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="IMAGE:7096948"  
 /tissue\_type="lung, pooled"  
 /lab\_host="DH10B T04"  
 /clone\_lib="NIH\_MGC\_231"  
 /notes="Organ: lung; Vector: pExpress-1; Site 1: EcoRV;  
 Site 2: NotI; RNA obtained from pooled lung tissue from a  
 mix of male and female animals at 8 wk old. Tissues were  
 snap-frozen and kept at -80C for two days before RNA  
 extraction and purification (Tri-reagent method). cDNA was  
 primed using oligo-dT primer:  
 5'-pGACTGTTCTAGTCGAGCGCGCCCTT)25-3' and cloned into  
 the EcoRV/NotI sites of pExpress-1. Size-selection >1.4kb  
 resulted in an average insert size of 2.3 kb. This primary  
 library is not normalized (normalized primary library is  
 NIH MGC 232) and was constructed by Express Genomics  
 (Frederick, MD). Note: this is a NIH\_MGC library."

#### ORIGIN

Query Match 82.9%; Score 17.4; DB 7; Length 721;  
 Best Local Similarity 94.7%; Pred. No. 9.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 636 GCATGTATGGTTATGTC 654

RESULT 38  
 CB245585  
 LOCUS  
 DEFINITION UI-M-FY0-cdt-o-22-0-UI.r1 NIH\_BMAP\_FY0 Mus musculus cDNA clone  
 IMAGE:6834383 5', mRNA sequence.

ACCESSION CB245585  
 VERSION CB245585.1 GI:28367229

KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 767)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)

#### FEATURES

1..767  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"

/db\_xref="taxon:10090"  
 /clone="IMAGE:6834383"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 13.5, 14.5, 16.5, 17.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH\_BMAP\_FY0"

/note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according  
 to Bonaudo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 is AGCGACACAG. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP). Gene Discovery in the  
 Developing Mouse Nervous System, supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

#### ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 767;  
 Best Local Similarity 94.7%; Pred. No. 9.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 292 GCATGTATGGTTATGTC 310

#### ORIGIN

RESULT 39  
 CO428850  
 LOCUS  
 DEFINITION UI-M-HX0-crw-m-17-0-UI.r1 NIH\_BMAP\_HX0 Mus musculus cDNA clone  
 IMAGE:30686416 5', mRNA sequence.

ACCESSION CO428850  
 VERSION CO428850.1 GI:49675144

KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 768)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 The following repetitive elements were found in this cDNA  
 sequence: 175-222, >(CAG)n#Simple\_repeat (matched complement)  
 Seq primer: pYX-5.

#### FEATURES

1..768  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30686416"  
 /tissue\_type="whole eye"  
 /dev\_stage="newborn (1, 5, 15 days) and embryonic (15, 16,  
 17, 18 dpc)"  
 /lab\_host="DH10B (T1 phage resistant)"

/clone lib="NIH\_BMAP\_HX0"  
 /notes="Organ: Eye; Vector: pYX-Asc; Site\_1: Ecor I;  
 Site\_2: Not I; The library was constructed according  
 Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AATAAATACG. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 7; Length 768;  
 Best Local Similarity 94.7%; Pred. No. 9.8e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 664 GCATGTATGGTTATCTTC 682

## RESULT 40

CD349045 777 bp mRNA linear EST 09-JUL-2003  
 LOCUS UI-M-FY0-cfp-g-14-0-UI.r1 NIH\_BMAP\_FY0 Mus musculus cDNA clone  
 DEFINITION IMAGE:6854151 5', mRNA sequence.

ACCESSION CD349045  
 VERSION 1 (bases 1 to 777)  
 KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 777)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

<http://genome.uiowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

## FEATURES

source

Location/Qualifiers

1..777  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6854151"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 13.5,14.5,16.5,17.5dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone lib="NIH\_BMAP\_FY0"

/notes="Organ: Brain; Vector: pYX-Asc; Site 1: Ecor I;

Site 2: Not I; The library was constructed according  
 Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned

directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is ACCGACAG. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 777;  
 Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGTTATCTTC 20  
 |||||  
 Db 23 GCATGTATGGTTATCTTC 41

## RESULT 41

CD349045 787 bp mRNA linear EST 12-JAN-2005  
 LOCUS UI-M-IB0-cue-d-03-0-UI.r1 NIH\_BMAP\_IB0 Mus musculus cDNA clone  
 DEFINITION IMAGE:30940682 5', mRNA sequence.

ACCESSION CD349045

VERSION 1 (bases 1 to 787)

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 787)

NIH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Dr. James Lin University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
<http://genome.uiowa.edu/distribution/mousefl.html>  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

## FEATURES

source

Location/Qualifiers

1..787  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:30940682"  
 /tissue\_type="whole eye"  
 /dev\_stage="newborn (1,5,15 days)"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone lib="NIH\_BMAP\_IB0"

/note="Organ: Eye; Vector: pYX-Asc; Site\_1: Ecor I;

Site\_2: Not I; The library was constructed according  
 Bonaïdo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with Ecor I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is AATAAATACG. This library was created for the University  
 of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 8; Length 787;  
 Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGATGGGTATCTTC 20  
 |||||  
 Db 293 GCATGATGGGTATGTTTC 311

## RESULT 42

CB521799  
 LOCUS  
 DEFINITION  
 UT-M-GHO-cen-n-08-0-UI.r1 NIH BMAP\_GHO Mus musculus cDNA clone  
 IMAGE:6842025 5', mRNA sequence.

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

CB521799 793 bp mRNA linear EST 09-JUL-2003  
 IMAGE:6842025 5', mRNA sequence.  
 CB521799 GI:29355154  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

1 (bases 1 to 793)  
 NIH-MGC http://mgc.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. Jim Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Distribution information can be found at  
 http://genome.uiowa.edu/distribution/mousefl.html  
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

The following repetitive elements were found in this cDNA  
 sequence: 83-130, >(CAG)nSimple\_repeat (matched complement)

Seq primer: pYX-5.

## FEATURES

source

Location/Qualifiers

1..793  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6842025"  
 /tissue\_type="whole brain"  
 /dev\_stage="1, 5, and 15 days newborn"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /clone\_lib="NIH BMAP\_GHO"  
 /notes="Organ: Brain; Vector: pYX-Asc; Site\_1: EcoR I;  
 Site\_2: Not I; The library was constructed according  
 Bernaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is CGACTGAGAT. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,  
 program coordinator."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 793;  
 Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGATGGGTATCTTC 20  
 |||||

Db 571 GCATGATGGGTATGTTTC 589

## RESULT 43

LOCUS

DEFINITION

CF737017

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

CF737017

EST.

Mus musculus

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.

1 (bases 1 to 797)

NIH-MGC http://mgc.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Dr. James Lin University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Distribution information can be found at

http://genome.uiowa.edu/distribution/mousefl.html

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1..797

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:30608409"

/tissue\_type="whole eye"

/dev\_stage="embryo 12.5, 13.5, 14.5 dpc"

/lab\_host="DH10B (T1 phage resistant)"

/clone\_lib="NIH BMAP\_HDO"

/notes="Organ: Eye; Vector: pYX-Asc; Site\_1: EcoR I;  
 Site\_2: Not I; The library was constructed according  
 Bernaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured RNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with oligo-dT  
 primer containing a Not I site. Double strand cDNA was  
 size selected according to mRNA size fraction, ligated  
 with EcoR I adaptor, digested with NotI and then cloned  
 directionally into pYX-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is TTATTGAAGT. This library was created for the University  
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
 Developing Mouse Nervous System', supported by National  
 Institute of Mental Health (NIMH)."

## ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 797;  
 Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGATGGGTATCTTC 20

|||||

Db 181 GCATGATGGGTATGTTTC 199

## RESULT 44

LOCUS

DEFINITION

CB290090

ACCESSION

CB290090

EST 09-JUL-2003

UI-M-F10-cdx-e-14-0-UI.r1 NIH BMAP\_F10 Mus musculus cDNA clone

IMAGE:6835671 5', mRNA sequence.



```

VERSION      CB290090.1  GI:28611505
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Dr. James Lin, University of Iowa
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              This clone was contributed by the Brain Molecular Anatomy Project
              (BMAP)

The following repetitive elements were found in this cDNA
sequence: 186-233, >(CAG)n#simple_repeat (matched complement)
Seq primer: pYX-5,

FEATURES
source
    Location/Qualifiers
        1..802
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE:6835671"
            /tissue_type="whole brain"
            /dev_stage="embryo 12.5dpc"
            /lab_host="DH10B (T1 phage resistant)"
            /clone_lib="NIH BMAP FTO"
            /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
            Site 2: Not I; The library was constructed according
            to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with EcoR I adaptor, digested with NotI and then cloned
            directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is CAGCCACGAC. This library was created for the University
            of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
            program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 802;
Best Local Similarity 94.7%; Pred. No. 9.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGTATGGGTATCTTC 20
        |||||
Db      674 GCATGTATGGGTATGTTTC 692

RESULT 45
LOCUS      BQ180689
DEFINITION      BQ180689
                UI-M-EX0-bxc-c-05-0-UI_r1 NIH BMAP_EX0 Mus musculus cDNA clone
                IMAGE:5706172 5', mRNA sequence.
ACCESSION      BQ180689
VERSION        BQ180689.1  GI:20356181
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Dr. Jim Lin, University of Iowa
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              This clone was contributed by the Brain Molecular Anatomy Project
              (BMAP)

The following repetitive elements were found in this cDNA
sequence: 186-233, >(CAG)n#simple_repeat (matched complement)
Seq primer: pYX-5,

FEATURES
source
    Location/Qualifiers
        1..802
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE:6835671"
            /tissue_type="whole brain"
            /dev_stage="embryo 12.5dpc"
            /lab_host="DH10B (T1 phage resistant)"
            /clone_lib="NIH BMAP FTO"
            /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
            Site 2: Not I; The library was constructed according
            to Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. Denatured RNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with oligo-dT
            primer containing a Not I site. Double strand cDNA was
            size selected according to mRNA size fraction, ligated
            with EcoR I adaptor, digested with NotI and then cloned
            directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA tail
            is CAGCCACGAC. This library was created for the University
            of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the
            Developing Mouse Nervous System', supported by National
            Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
            program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 6; Length 802;
Best Local Similarity 94.7%; Pred. No. 9.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGTATGGGTATCTTC 20
        |||||
Db      674 GCATGTATGGGTATGTTTC 692

RESULT 46
LOCUS      CF726722
DEFINITION      CF726722
                UI-M-HB0-cki-i-01-0-UI_r1 NIH BMAP_HB0 Mus musculus cDNA clone
                IMAGE:30548448 5', mRNA sequence.
ACCESSION      CF726722
VERSION        CF726722.1  GI:37600890
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Dr. James Lin, University of Iowa
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              This clone was contributed by the Brain Molecular Anatomy Project
              (BMAP)

The following repetitive elements were found in this cDNA
sequence: 251-298, >(CAG)n#simple_repeat (matched complement)
Seq primer: pYX-5,

FEATURES
source
    Location/Qualifiers
        1..804
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6"
            /db_xref="taxon:10090"
            /clone="IMAGE:5706172"
            /tissue_type="whole brain"
            /dev_stage="embryo 15.5 dpc"
            /lab_host="DH10B (T1 phage resistant)"
            /clone_lib="NIH BMAP EX0"
            /note="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I;
            Site 2: Not I; The library was constructed according to
            Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
            1996. Denatured mRNA was size fractionated on a 1% agarose
            gel. First strand cDNA synthesis was primed with an
            oligo-dT primer containing a Not I site. Double stranded
            cDNA was size selected according to mRNA size fraction,
            ligated with EcoR I adaptor, digested with Not I, and then
            cloned directionally into pYX-Asc vector. The library tag
            sequence located between the Not I site and the polyA
            tail, is GTGCGTGGA. This library was created for the
            University of Iowa Mouse Brain Molecular Anatomy Project
            (BMAP). 'Gene Discovery in the Developing Mouse Nervous
            System', supported by National Institutes of Mental Health
            (NIMH), Hemin Chin, Ph.D., program coordinator."

ORIGIN
Query Match      82.9%; Score 17.4; DB 3; Length 804;
Best Local Similarity 94.7%; Pred. No. 9.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GCATGTATGGGTATCTTC 20
        |||||
Db      739 GCATGTATGGGTATGTTTC 757

RESULT 46
LOCUS      CF726722
DEFINITION      CF726722
                UI-M-HB0-cki-i-01-0-UI_r1 NIH BMAP_HB0 Mus musculus cDNA clone
                IMAGE:30548448 5', mRNA sequence.
ACCESSION      CF726722
VERSION        CF726722.1  GI:37600890
KEYWORDS
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Dr. James Lin, University of Iowa
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              This clone was contributed by the Brain Molecular Anatomy Project
              (BMAP)

The following repetitive elements were found in this cDNA
sequence: 251-298, >(CAG)n#simple_repeat (matched complement)
Seq primer: pYX-5,

```



JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mousefl.html  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)  
The following repetitive elements were found in this cDNA  
sequence: 96-143, >(CAG)n\$imple\_repeat (matched complement)  
Seq primer: pYX-5.

## FEATURES

source

Location/Qualifiers

```
1. .806
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:30548448"
/tissue_type="whole eye"
/dev_stage="embryo 12.5,13.5,14.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP HB0"
/notes="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TTATGTAGT. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH)."
```

## ORIGIN

Query Match 82.9%; Score 17.4; DB 6; Length 806;  
Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTTATCTTC 20

```
|||||
584 GCATGTATGGTTATGTC 602
```

RESULT 47

BQ444317

LOCUS

DEFINITION BQ444317 UI-M-EXO-bxk-k-22-0-UI-r1 NIH BMAP\_EXO Mus musculus cDNA clone EST 29-MAY-2002  
IMAGE:5709453 5', mRNA sequence.

ACCESSION

BQ444317

VERSION

BQ444317.1

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 814)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa

cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

Seq primer: pYX-5.

## FEATURES

source

Location/Qualifiers

```
1. .814
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5709453"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP EX0"
/notes="Organ: Brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is GTGGGTGGA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
```

## ORIGIN

Query Match 82.9%; Score 17.4; DB 5; Length 814;  
Best Local Similarity 94.7%; Pred. No. 9.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTTATCTTC 20

```
|||||
46 GCATGTATGGTTATGTC 64
```

RESULT 48

B1685030

LOCUS

DEFINITION B1685030 603310009F1 NCI\_CGAP\_Mam6 Mus musculus cDNA clone IMAGE:5345930 5', mRNA sequence.

ACCESSION

B1685030

VERSION

B1685030.1

KEYWORDS

EST.

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 840)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Jeffrey Green M.D.

cDNA Library preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: L1AM11879 row: a column: 03

High quality sequence stop: 813.

Location/Qualifiers

1. .840

source

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5345930"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI CGAP Mam6"
/notes="Organ: mammary; Vector: pCMV-SPORT6; Site 1: Salt; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

```

## ORIGIN

```

Query Match      82.9%; Score 17.4; DB 3; Length 840;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 GCATGATGGGTATCTTC 20
    |||||
DB  739 GCATGATGGGTATCTTC 757

```

## RESULT 49

```

LOCUS      BU613524
DEFINITION UI-M-EWO-cas-p-10-0-UI.r1 NIH BMAP EWO Mus musculus cDNA clone
ACCESSION  BU613524
VERSION    BU613524.1 GI:23279739
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
           Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 845)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: gcapbs-remail.nih.gov
           Tissue Procurement: Dr. James Lin, University of Iowa
           cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
           cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
           DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
           Clone Distribution: Clone distribution information can be obtained
           from Dr. M. Bento Soares, bentso-soares@uiowa.edu
           This clone was contributed by the Brain Molecular Anatomy Project
           (BMAP)
Seq primer: pyX-5.

```

## FEATURES

```

source
1..845
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="UI-M-EWO-cas-p-10-0-UI"
/tissue_type="whole brain"
/dev_stage="embryo 15.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP EWO"
/notes="Organ: brain; Vector: pyX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to
Ronald, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pyX-Asc vector. The library tag

```

sequence located between the Not I site and the polyA tail, is GTGCGTGA. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (EMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

## ORIGIN

```

Query Match      82.9%; Score 17.4; DB 5; Length 845;
Best Local Similarity 94.7%; Pred. No. 1e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  2 GCATGATGGGTATCTTC 20
    |||||
DB  441 GCATGATGGGTATCTTC 459

```

## RESULT 50

```

LOCUS      CNS0460Q/c
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
ACCESSION  AL277019.1 GI:8011222
VERSION    AL277019
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetraodon nigroviridis
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
           Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
           Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE  1
AUTHORS   Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
           Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
           Saurin,W. and Weissenbach,J.
TITLE     Estimate of human gene number provided by genome-wide analysis
           using Tetraodon nigroviridis DNA sequence
JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
PUBMED    10835645
REFERENCE  2
AUTHORS   Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
           Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
           Saurin,W., Bernot,A. and Weissenbach,J.
TITLE     Characterization and repeat analysis of the compact genome of the
           freshwater pufferfish Tetraodon nigroviridis
JOURNAL   Genome Res. 10 (7), 939-949 (2000)
PUBMED    10899143
REFERENCE  3 (bases 1 to 961)
AUTHORS   Genoscope.
TITLE     Direct Submission
JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
           BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
           - Web : www.genoscope.cns.fr)
COMMENT   This sequence is a single read and was generated as part of a large
           scale clone-end sequencing project of the Tetraodon nigroviridis
           genome. For more information, please take a look at
           http://www.genoscope.cns.fr/Tetraodon.

```

## FEATURES

```

source
1..961
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="086K03"
/clone_lib="G"
/notes="Genoscope sequence ID : COBG086AF02LP1
end : T7"

```

## ORIGIN

```

Query Match      82.9%; Score 17.4; DB 11; Length 961;
Best Local Similarity 85.7%; Pred. No. 1e+03;
Matches 18; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 TGCATGTATGGTTATCTCC 21  
||||| ||||| ||||| ||||| |||||  
Db 623 TGCATGTCTGGTTTCTYCC 603

RESULT 51  
BG519143 995 bp mRNA linear EST 02-APR-2001  
DEFINITION 602578081F1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:3481653 5',  
mRNA sequence.  
ACCESSION BG519143  
VERSION BG519143.1 GI:13514654  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 995)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-x@mail.nih.gov](mailto:cgapbs-x@mail.nih.gov)  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM8510 row: b column: 22  
High quality sequence stop: 741.  
Location/Qualifiers

#### FEATURES

source  
1..995  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="mix FVB/N, C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3481653"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/clone\_lib="NCI\_CGAP\_Mam5"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

#### ORIGIN

Query Match 82.9%; Score 17.4; DB 2; Length 995;  
Best Local Similarity 94.7%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GCATGTATGGTTATCTTC 20  
||||| ||||| ||||| ||||| |||||  
Db 326 GCATGTATGGTTATGTTTC 344

#### RESULT 52

CL116220 1032 bp DNA linear GSS 05-JAN-2004  
DEFINITION ISB1-63M20 T7.1 ISB1 Xenopus tropicalis genomic clone ISB1-63M20,  
genomic survey sequence.  
ACCESSION CL116220  
VERSION CL116220.1 GI:40609855  
KEYWORDS GSS.  
SOURCE Xenopus tropicalis (western clawed frog)  
ORGANISM Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

#### REFERENCE

AUTHORS Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,  
Mardis, E. and Wilson, R.  
TITLE A physical map of the xenopus tropicalis genome  
JOURNAL Unpublished (2003)  
COMMENT Contact: Richard K Wilson  
Genome Sequencing Center  
Washington University School of Medicine  
Email: [submissions@wustl.edu](mailto:submissions@wustl.edu)  
Insert Length: 75000 Std Error: 0.00  
Seq primer: T7 TAATACGACTCACTATAGGG  
Class: BAC ends  
High quality sequence start: 8  
High quality sequence stop: 374.  
Location/Qualifiers

#### FEATURES

source  
1..1032  
/organism="Xenopus tropicalis"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:8364"  
/clone="ISB1-63M20"  
/clone\_lib="ISB1"  
/note="Vector: pBelOBAC11; ISB1 Xenopus tropicalis BAC  
Library Segment 1"  
Library Segment 1"

#### ORIGIN

Query Match 82.9%; Score 17.4; DB 10; Length 1032;  
Best Local Similarity 94.7%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 TGCATGTATGGTTATCTTC 19  
||||| ||||| ||||| ||||| |||||  
Db 40 TGCATTATGGTTATCTTC 58

#### RESULT 53

AK029006 3610 bp mRNA linear HTC 03-APR-2004  
LOCUS Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched  
DEFINITION library, clone:4732480N13 product:DnaJ (Hsp40) homolog, subfamily  
C, member 5, full insert sequence.  
ACCESSION AK029006  
VERSION AK029006.1 GI:26324963  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

#### REFERENCE

AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
PUBMED 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitagawa, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuda, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multipillar sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
PUBMED 11076861  
REFERENCE 4

AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.			RESULT 54 AK032373 LOCUS DEFINITION	AK032373 3626 bp mRNA linear HTC 03-APR-2004 Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430529B06 product:DnaJ (Hsp40) homolog, subfamily C, member 5, full insert sequence.			
	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.				AK032373.1 GI:26328200 HTC; CAP trapper.			
	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				Mus musculus (house mouse)			
JOURNAL REFERENCE AUTHORS	Nature 420, 563-573 (2002)			ACCESSION VERSION KEYWORDS SOURCE ORGANISM	AK032373			
	6 (bases 1 to 3610)				Mus musculus			
	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.				Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
JOURNAL REFERENCE AUTHORS	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)			REFERENCE	Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)			
	CNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.				10349636			
	Please visit our web site for further details. URL: http://genome.gsc.riken.jp/ URL: http://fantom.gsc.riken.jp/ Location/Qualifiers				2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)			
FEATURES source	1. .3610 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM DB:4732480N13" /db_xref="taxon:10090" /clone="4732480N13" /tissue_type="skin" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 days neonate" 163. .759 /codon_start=1 /protein_id="BAC26236.1" /db_xref="GI:26324964" /translations="MADQRQSLSTGESLYHLGLDKNATSDDIKKSKYRKLAKYHP DKNPDNPEADKFEINNAHILTDATKNIYDKYGLYVAEQGEENVNTYFVLS SWAKALFVVGGLTCCYCCCLCCFCNCCGCKPKAPGEETEFVSPDELEAQLQ SDREATDTFIVQPASATTTQLTADSHFSYHTDGN"			TITLE JOURNAL PUBMED REFERENCE AUTHORS	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Hara, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)			
	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)				4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)			
	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)				5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002)			
CDS	6 (bases 1 to 3626) Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.			TITLE JOURNAL PUBMED REFERENCE AUTHORS	6 (bases 1 to 3626) Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.			
	Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216)				Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/, tel: 81-45-503-9222, Fax: 81-45-503-9216)			
	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.				cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken.			
ORIGIN	Query Match 82.9%; Score 17.4; DB 4; Length 3610; Best Local Similarity 94.7%; Pred. No. 1.2e+03; Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			Qy Db	2 GCATGTATGGTATCTTC 20       1001 GCATGTATGGTATCTTC 1019 			

Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.  
Please visit our web site for further details.  
URL: <http://genome.gsc.riken.jp/>  
URL: <http://fantom.gsc.riken.jp/>  
Location/Qualifiers

## FEATURES

source  
1. .3626  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:6430529806"  
/db\_xref="taxon:10090"  
/clone="6430529806"  
/sex="male"  
/tissue\_type="olfactory brain"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
157. .753  
CDS  
/note="unnamed protein product; DnaJ (Hsp40) homolog, subfamily C, member 5 (MGD|MGJ:892995, GB|NM\_016775, evidence: BLASTN, 99%, match=597)  
putative"  
/codon\_start=1  
/protein\_id="BAC27841.1"  
/db\_xref="GI:26328201"  
/translation="MADQRORSLSSTGESLYHVLGDKNATSDDIKKSYRKLAKYHP  
DKNPDPAAADKFEINNAHALITDAIKRINYKYSGLGYVAEQGEENVYFVLS  
SWWAKALFVVCGLLTCCYCCCCGKCKPKAPGSETEFVSPDELEAQLQ  
SDREATDTPIVIQPASATETQLTADSPHSYHTDGFN"

## ORIGIN

Query Match 82.9%; Score 17.4; DB 4; Length 3626;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 GCATGTATGGTATCTTC 20  
|||||  
Db 995 GCATGTATGGTATGTTTC 1013  
|||||

## RESULT 55

AK083584  
LOCUS  
DEFINITION  
Mus musculus 9 days embryo whole body cDNA, RIKEN full-length enriched library, clone:D030049H18 product:DnaJ (Hsp40) homolog, subfamily C, member 5, full insert sequence.

## ACCESSION

VERSION  
AK083584.1 GI:26101348

## KEYWORDS

SOURCE  
Mus musculus (house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
10349636  
2  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, N., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
11042159  
3  
Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, N., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## PUBMED

## REFERENCE

## AUTHORS

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
11076861

## 4

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

## 5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

## 6

(bases 1 to 3732)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: <http://genome.gsc.riken.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.jp/>

URL: <http://fantom.gsc.riken.jp/>

Location/Qualifiers

1. .3732

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="FANTOM DB:D030049H18"

/db\_xref="taxon:10090"

/clone="D030049H18"

/tissue\_type="whole body"

/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev\_stage="9 days embryo"

1. .3732

/notes="DnaJ (Hsp40) homolog, subfamily C, member 5 (MGD|MGJ:892995, GB|NM\_016775, evidence: BLASTN, 99%, match=597)"

misc\_feature

Query Match 82.9%; Score 17.4; DB 4; Length 3732;  
Best Local Similarity 94.7%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GCATGTATGGTATCTTC 20  
|||||

Db 1100 GCATGTATGGTATGTTTC 1118  
|||||

RESULT 56

CG735819

```

LOCUS       CG735819               668 bp    DNA        linear        GSS 22-OCT-2003
DEFINITION  ZMMBB0321B11.f ZMMBB Zea mays genomic clone ZMMBB0321B11 5',
genomic survey sequence.
ACCESSION   CG735819
VERSION     CG735819.1  GI:37808544
KEYWORDS    GSS.
SOURCE      Zea mays
ORGANISM    Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 668)
AUTHORS     Yu, Y., Kim, H.R., Hatfield, J., Soderlund, C., Bharti, A.K., Messing, J.
            and Wing, R.K.
TITLE       Sequencing of the maize genome
JOURNAL     Unpublished (2003)
COMMENT     Contact: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: rwing@genome.arizona.edu
PCR PRIMERS
FORWARD: T7
BACKWARD: M13r
Plate: 0321 row: B column: 11
Seq primer: T7
Class: BAC ends.

FEATURES             Location/Qualifiers
     source           1..668
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBB0321B11"
                     /lab_host="DH10B"
                     /clone_lib="ZMMBB"
                     /notes="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
                     HindIII; Zea mays L. ssp. mays"

ORIGIN
Query Match      81.0%; Score 17; DB 10; Length 668;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  TGCATGTATGGGTTATC 17
    |||||||
DB  174 TGCATGTATGGGTTATC 190

RESULT 57
CN212867/c
LOCUS       CN212867               738 bp    mRNA        linear        EST 08-APR-2004
DEFINITION  28396 Suspension culture Solanum tuberosum cDNA, mRNA sequence.
ACCESSION   CN212867
VERSION     CN212867.1  GI:46292209
KEYWORDS    EST.
SOURCE      Solanum tuberosum (potato)
ORGANISM    Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
            asterids; lamiales; Solanales; Solanaceae; Solanum.
REFERENCE   1 (bases 1 to 738)
AUTHORS     Flinn, B., Rothwell, C., Sardana, R., Griffiths, R., Laque, M., De
            Koeper, D., Audy, P., Goyer, C., Li, X.-Q., Wang, Pruski, G. and Regan, S.
            Generation of ESTs from potato suspension cultures
            Unpublished (2004)
TITLE       The Canadian Potato Genome Project - BioAtlantech
JOURNAL     921 College Hill Rd, Fredericton, ON, E3B 6Z9, CANADA
COMMENT     Email: bflinn@bioatlantech.nb.ca

LOCUS       BW810341               109 bp    mRNA        linear        EST 23-MAY-2005
DEFINITION  BW810341 Amphioxus Branchiostoma floridae unpublished cDNA library,
larva whole animal Branchiostoma floridae cDNA clone bslv047c21 5',
mRNA sequence.
ACCESSION   BW810341
VERSION     BW810341.1  GI:66418557
KEYWORDS    EST.
SOURCE      Branchiostoma floridae (Florida lancelet)
ORGANISM    Branchiostoma floridae
            Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
            Branchiostoma.
REFERENCE   1 (bases 1 to 109)
AUTHORS     Yu, J., Holland, L.Z., Shin-i, T., Kohara, Y., Satou, Y. and Satoh, N.
TITLE       Expressed genes in Branchiostoma floridae
JOURNAL     Unpublished (2005)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.

FEATURES             Location/Qualifiers
     source           1..109
                     /organism="Branchiostoma floridae"
                     /mol_type="mRNA"
                     /db_xref="taxon:7739"
                     /clone="bslv047c21"
                     /tissue_type="whole animal"
                     /dev_stage="larva"
                     /clone_lib="Amphioxus Branchiostoma floridae unpublished
                     cDNA library, larva whole animal"

ORIGIN
Query Match      80.0%; Score 16.8; DB 5; Length 109;
Best Local Similarity 90.0%; Pred. No. 1.5e+03;

```



Job time : 1712.94 secs

*This Page Blank (uspto)*

---



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:56:54 ; Search time 58.9888 Seconds  
(without alignments)  
632.812 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcattatgggtattcttcc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 60 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
- 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
- 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
- 6: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq:\*
- 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
- 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
- 9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	100.0	21	3	US-10-081-923-2
2	21	100.0	1323	3	US-10-081-923-12
3	21	100.0	1378	3	US-10-081-923-14
4	21	100.0	1379	3	US-10-081-923-7
5	21	100.0	1379	3	US-10-081-923-13
6	21	100.0	1379	3	US-10-081-923-16
7	21	100.0	1384	3	US-10-081-923-8
8	21	100.0	1384	3	US-10-081-923-10
9	21	100.0	1385	3	US-10-081-923-9
10	21	100.0	1390	3	US-10-081-923-11
11	21	100.0	1393	3	US-10-081-923-15
12	21	100.0	1447	3	US-10-081-923-17
13	21	100.0	1803	3	US-10-081-923-5
14	19.4	92.4	1140	3	US-09-583-110-1191
15	19.4	92.4	1287	3	US-09-107-433-1469
16	19.4	92.4	8195	3	US-08-961-527-94
17	16.8	80.0	89210	3	US-08-949-016-16988
18	16.8	80.0	144922	3	US-09-949-016-15890
19	16.2	77.1	1059	3	US-09-328-352-540
20	16	76.2	1356	3	US-09-248-796A-2116
21	16	76.2	4291	3	US-09-351-200-1
22	15.8	75.2	601	3	US-09-949-016-24376
23	15.8	75.2	601	3	US-09-949-016-24696
24	15.8	75.2	601	3	US-09-949-016-111355

```

25 15.8 75.2 601 3 US-09-949-016-160998 Sequence 160998,
26 15.8 75.2 601 3 US-09-949-016-199547 Sequence 199547,
27 15.8 75.2 8351 2 US-08-198-446B-14 Sequence 14, Appl
28 15.8 75.2 8351 2 US-08-870-693-14 Sequence 14, Appl
29 15.8 75.2 9721 3 US-09-949-016-14814 Sequence 14814, A
30 15.8 75.2 9721 3 US-09-949-016-12076 Sequence 12076, A
31 15.8 75.2 26831 3 US-09-949-016-16250 Sequence 16250, A
32 15.8 75.2 36755 3 US-09-949-016-16994 Sequence 16994, A
33 15.8 75.2 38059 3 US-09-328-925-4 Sequence 4, Appl
34 15.8 75.2 41435 3 US-09-949-016-15926 Sequence 15926, A
35 15.8 75.2 152132 3 US-09-949-016-13845 Sequence 13845, A
36 15.8 75.2 152145 3 US-09-949-016-12371 Sequence 12371, A
37 15.8 75.2 373182 3 US-09-949-016-12371 Sequence 12371, A
38 15.8 75.2 373694 3 US-09-949-016-12062 Sequence 12062, A
39 15.8 75.2 678533 3 US-09-949-016-14577 Sequence 14577, A
40 15.8 75.2 678533 3 US-09-949-016-14578 Sequence 14578, A
41 15.8 75.2 784019 3 US-09-949-016-14033 Sequence 14033, A
42 15.8 75.2 828152 3 US-09-949-016-12777 Sequence 12777, A
43 15.4 73.3 1965 3 US-09-948-796A-5759 Sequence 5759, Ap
44 15.4 73.3 91772 3 US-09-949-016-15568 Sequence 15568, A
45 15.4 73.3 106199 3 US-09-949-016-12393 Sequence 12393, A
46 15.4 73.3 107980 3 US-09-949-016-14370 Sequence 14370, A
47 15.4 73.3 205183 3 US-09-949-016-17009 Sequence 17009, A
48 15.2 72.4 573 2 US-08-286-819A-20 Sequence 20, Appl
49 15.2 72.4 573 3 US-08-980-357-20 Sequence 20, Appl
50 15.2 72.4 885 3 US-09-270-767-28376 Sequence 28376, A
51 15.2 72.4 885 3 US-09-270-767-12594 Sequence 12594, A
52 15.2 72.4 1306 3 US-08-286-819A-15 Sequence 15, Appl
53 15.2 72.4 7225 3 US-08-980-357-15 Sequence 15, Appl
54 15.2 72.4 7225 3 US-09-357-375-15 Sequence 15, Appl
55 15.2 72.4 7225 3 US-08-286-819A-16 Sequence 16, Appl
56 15.2 72.4 10851 2 US-08-980-357-16 Sequence 16, Appl
57 15.2 72.4 10851 3 US-09-357-375-16 Sequence 16, Appl
58 15.2 72.4 10851 3 US-09-357-375-16 Sequence 16, Appl
59 15.2 72.4 29121 3 US-09-949-016-11748 Sequence 11748, A
60 15.2 72.4 29122 3 US-09-949-016-13591 Sequence 13591, A

```

#### ALIGNMENTS

```

RESULT 1
US-10-081-923-2
; Sequence 2, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-10-081-923-2

```

```

Query Match 100.0%; Score 21; DB 3; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```
Oy 1 TGCATGTATGGGTATCTTCC 21
|||||

```

```
Db 1 TGCATGTATGGGTATCTTCC 21
|||||

```

RESULT 2

US-10-081-923-12/c

```
; Sequence 12, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 2
US-10-081-923-12

Query Match          100.0%; Score 21; DB 3; Length 1323;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   ||||||||||||||||||
Db 193 TGCATGTATGGGTATCTTCC 173

RESULT 3
US-10-081-923-14/c
; Sequence 14, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1378
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 10
US-10-081-923-14

Query Match          100.0%; Score 21; DB 3; Length 1378;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   ||||||||||||||||||
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 4
US-10-081-923-7/c
; Sequence 7, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 1
US-10-081-923-7

Query Match          100.0%; Score 21; DB 3; Length 1379;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   ||||||||||||||||||
Db 210 TGCATGTATGGGTATCTTCC 190

; Sequence 12, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 2
US-10-081-923-12

Query Match          100.0%; Score 21; DB 3; Length 1323;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   ||||||||||||||||||
Db 193 TGCATGTATGGGTATCTTCC 173

RESULT 3
US-10-081-923-14/c
; Sequence 14, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1378
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 10
US-10-081-923-14

Query Match          100.0%; Score 21; DB 3; Length 1378;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   ||||||||||||||||||
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 4
US-10-081-923-7/c
; Sequence 7, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 1
US-10-081-923-7

Query Match          100.0%; Score 21; DB 3; Length 1379;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
   ||||||||||||||||||
Db 210 TGCATGTATGGGTATCTTCC 190
```

```
RESULT 7
US-10-081-923-8/c
; Sequence 8, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 5
US-10-081-923-8
Query Match 100.0%; Score 21; DB 3; Length 1384;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 8
US-10-081-923-10/c
; Sequence 10, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 8
US-10-081-923-10
Query Match 100.0%; Score 21; DB 3; Length 1384;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 9
US-10-081-923-9/c
; Sequence 9, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 7
US-10-081-923-9
Query Match 100.0%; Score 21; DB 3; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
Db 210 TGCATGTATGGTTATCTTCC 190

RESULT 10
US-10-081-923-11/c
; Sequence 11, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 9
US-10-081-923-11
Query Match 100.0%; Score 21; DB 3; Length 1390;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21
Db 224 TGCATGTATGGTTATCTTCC 204

RESULT 11
US-10-081-923-15/c
; Sequence 15, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pISI sequence from isolate no. 3
US-10-081-923-15
Query Match 100.0%; Score 21; DB 3; Length 1393;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 TGCATGTATGGGTATCTTCC 21  
Db 210 TGCATGTATGGGTATCTTCC 190

## RESULT 12

US-10-081-923-17/c  
; Sequence 17, Application US/10081923  
; Patent No. 6593093  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/081,923  
; CURRENT FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17  
; LENGTH: 1447  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: ptsI sequence from isolate no. 11  
US-10-081-923-17

Query Match 100.0%; Score 21; DB 3; Length 1447;  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
Db 210 TGCATGTATGGGTATCTTCC 190

## RESULT 13

US-10-081-923-5/c  
; Sequence 5, Application US/10081923  
; Patent No. 6593093  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/081,923  
; CURRENT FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1803  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: ptsI sequence from Oklahoma University M1 strain  
; PUBLICATION INFORMATION:  
; AUTHORS: Ferretti et al.  
; JOURNAL: Proc. Natl. Acad. Sci. USA  
; VOLUME: 98  
; PAGES: 4658-4663  
; DATE: 2001-01-01  
US-10-081-923-5

Query Match 100.0%; Score 21; DB 3; Length 1803;  
Best Local Similarity 100.0%; Pred. No. 0.57;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
Db 377 TGCATGTATGGGTATCTTCC 357

## RESULT 14

US-09-583-110-1191/c  
; Sequence 1191, Application US/09583110  
; Patent No. 669703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; TITLE OF INVENTION: Pneumoniae for Diagnostics and Therapeutics  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 1191  
; LENGTH: 1140  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-1191

Query Match 92.4%; Score 19.4; DB 3; Length 1140;  
Best Local Similarity 95.2%; Pred. No. 3.4;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
Db 377 TGCATGTATGGGTATCTTCC 357

RESULT 15

US-09-107-433-1469/c  
; Sequence 1469, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 1469:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1287 base pairs  
; TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pneumoniae  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...1287  
SEQUENCE DESCRIPTION: SEQ ID NO: 1469:  
US-09-107-433-1469

Query Match 92.4%; Score 19.4; DB 3; Length 1387;  
Best Local Similarity 95.2%; Pred. No. 3.4; 1; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTC 21  
|||||  
Db 431 TGCATGTATGGTTATCTTC 411

RESULT 16  
US-08-961-527-94/c  
; Sequence 94, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8195 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

Query Match 92.4%; Score 19.4; DB 3; Length 8195;  
Best Local Similarity 95.2%; Pred. No. 4.6; 1; Indels 0; Gaps 0;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTC 21  
|||||  
Db 1384 TGCATGTATGGTTATCTTC 1364

RESULT 17  
US-09-949-016-16988  
; Sequence 16988, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16988  
; LENGTH: 89210  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16988

Query Match 80.0%; Score 16.8; DB 3; Length 89210;  
Best Local Similarity 90.0%; Pred. No. 1.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTC 20  
|||||  
Db 70426 TGCATGTATGGTTATCTTC 70445

RESULT 18  
US-09-949-016-15890  
; Sequence 15890, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15890  
; LENGTH: 144922  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15890

Query Match 80.0%; Score 16.8; DB 3; Length 144922;  
Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTC 20  
|||||  
Db 130653 TGCATGTATGGTTATCTTC 130672

RESULT 19  
US-09-328-352-540  
; Sequence 540, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:

```
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 540
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-540

Query Match          77.1%; Score 16.2; DB 3; Length 1059;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21
    ||||| ||||| ||||| ||||| |||||
Db 561 TGCATTTATGGCTTATTTCC 581

RESULT 20
US-09-248-796A-2116/c
; Sequence 2116, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 2116
; LENGTH: 1356
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-2116

Query Match          76.2%; Score 16; DB 3; Length 1356;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGTATGGGTTATCTTC 20
    ||||| ||||| ||||| |||||
Db 1258 TGTATGGGTTATCTTC 1243

RESULT 21
US-09-351-200-1/c
; Sequence 1, Application US/09351200
; Patent No. 6320033
; GENERAL INFORMATION:
; APPLICANT: BOURBONNAIS, Yves
; APPLICANT: LAMARE, Claude
; APPLICANT: DESLAURIERS, No. 63200331a
; TITLE OF INVENTION: CANDIDA ALBICANS GENE (CSA1) ENCODING A
; FILE REFERENCE: 6013-71"US" CC/
; CURRENT APPLICATION NUMBER: US/09/351,200
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: CA2,237,134
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 4291
; TYPE: DNA
```

```
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: gene
; LOCATION: (0)...(0)
US-09-351-200-1

Query Match          76.2%; Score 16; DB 3; Length 4291;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGTATGGGTTATCTTC 20
    ||||| ||||| ||||| |||||
Db 2186 TGTATGGGTTATCTTC 2171

RESULT 22
US-09-949-016-24376/c
; Sequence 24376, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24376
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24376

Query Match          75.2%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCCATGTATGGGTTATCTT 19
    ||||| ||||| ||||| |||||
Db 28 TGCATATATGAGTTATCTT 10

RESULT 23
US-09-949-016-24696
; Sequence 24696, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24696
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-24696
```

Query Match 75.2%; Score 15.8; DB 3; Length 601;  
Best Local Similarity 89.5%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Gaps 0;

Qy 3 CATGTATGGGTTATCTTC 21  
Db 128 CATGTATGGGTTAACTTC 146

RESULT 24  
US-09-949-016-111355  
; Sequence 111355, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 111355  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-111355

Query Match 75.2%; Score 15.8; DB 3; Length 601;  
Best Local Similarity 89.5%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Gaps 0;

Qy 3 CATGTATGGGTTATCTTC 21  
Db 128 CATGTATGGGTTAACTTC 146

RESULT 25  
US-09-949-016-160998  
; Sequence 160998, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 160998  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-160998

Query Match 75.2%; Score 15.8; DB 3; Length 601;  
Best Local Similarity 89.5%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTT 19  
Db 138 TGCATGTATGGGTTACCAT 156

RESULT 26  
US-09-949-016-199547/c  
; Sequence 199547, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 199547  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-199547

Query Match 75.2%; Score 15.8; DB 3; Length 601;  
Best Local Similarity 89.5%; Pred. No. 2e+02; 2; Indels 0; Gaps 0;  
Matches 17; Conservative 0; Mismatches 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTT 19  
Db 28 TGCATATATGAGTTATCTT 10

RESULT 27  
US-08-198-446B-14/c  
; Sequence 14, Application US/08198446B  
; Patent No. 5674996  
; GENERAL INFORMATION:  
; APPLICANT: Hartwell, Leland H.  
; APPLICANT: Weinert, Ted A.  
; APPLICANT: Plon, Sharon E.  
; APPLICANT: Groudine, Mark T.  
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC  
; STREET: 1420 Fifth Ave., Suite 2800  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98101-2347  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/198,446B  
; FILING DATE: 18-FEB-1994  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sheiness, Diana K.  
; REGISTRATION NUMBER: 35,356  
; REFERENCE/DOCKET NUMBER: FHC17537  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-682-8100  
; TELEFAX: 206-224-0779

```
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; DESCRIPTION: yeast MEC1 CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 784..7890
; US-08-198-4468-14

Query Match 75.2%; Score 15.8; DB 2; Length 8351;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTATCTT 19
||||| ||| ||||| |||||
DB 7058 TGCATATATGGTATCTT 7040

RESULT 28
US-08-870-693-14/c
; Sequence 14, Application US/08870693
; Patent No. 5866338
; GENERAL INFORMATION:
; APPLICANT: Hartwell, Leland H.
; APPLICANT: Weinert, Ted A.
; APPLICANT: Plon, Sharon E.
; APPLICANT: Groudine, Mark T.
; TITLE OF INVENTION: Cell Cycle Checkpoint Genes
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Ave., Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,693
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/198,446
; FILING DATE: February 18, 1994
; APPLICATION NUMBER: PCT/US93/04458
; FILING DATE: May 12, 1993
; APPLICATION NUMBER: US 07/884,426
; FILING DATE: May 14, 1992
; APPLICATION NUMBER: US 07/882,051
; FILING DATE: May 12, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR110798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100
; TELEFAX: 206-224-0779
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; DESCRIPTION: yeast MEC1 CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 784..7890
; US-08-870-693-14

Query Match 75.2%; Score 15.8; DB 2; Length 8351;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTATCTT 19
||||| ||| ||||| |||||
DB 7058 TGCATATATGGTATCTT 7040

RESULT 29
US-09-949-016-14814
; Sequence 14814, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14814
; LENGTH: 9721
; TYPE: DNA
; ORGANISM: Human
; US-09-949-016-14814

Query Match 75.2%; Score 15.8; DB 3; Length 9721;
Best Local Similarity 89.5%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTATCTTCC 21
||||| ||||| |||||
DB 2756 CATGTATGGGTAACTTTC 2774

RESULT 30
US-09-949-016-12076
; Sequence 12076, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```



```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12076
; LENGTH: 9722
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12076

Query Match          75.2%; Score 15.8; DB 3; Length 9722;
Best Local Similarity 89.5%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 CATGTATGGGTTATCTCC 21
      ||||| ||||| ||||| |||||
Db      2761 CATGTATGGGTTAACTTC 2779

RESULT 31
US-09-949-016-16250/c
; Sequence 16250, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16250
; LENGTH: 26831
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16250

Query Match          75.2%; Score 15.8; DB 3; Length 26831;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
      ||||| ||||| ||||| |||||
Db      2875 TGCATGTATGGGTTACCAT 2857

RESULT 32
US-09-949-016-16994/c
; Sequence 16994, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16994
; LENGTH: 36755
; TYPE: DNA

; ORGANISM: Human
US-09-949-016-16994

Query Match          75.2%; Score 15.8; DB 3; Length 36755;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
      ||||| ||||| ||||| |||||
Db      25570 TGCATGTATGAGTGATCTT 25552

RESULT 33
US-09-328-925-4/c
; Sequence 4, Application US/09328925
; Patent No. 6610906
; GENERAL INFORMATION:
; APPLICANT: Kurachi, Kotoku
; APPLICANT: Kurachi, Sumiko
; TITLE OF INVENTION: Nucleotide Sequences for Gene Regulation and Methods of
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: UM-03603
; CURRENT APPLICATION NUMBER: US/09/328,925
; CURRENT FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 38059
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-328-925-4

Query Match          75.2%; Score 15.8; DB 3; Length 38059;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
      ||||| ||||| ||||| |||||
Db      26539 TGCATGTATGAGTGATCTT 26521

RESULT 34
US-09-949-016-15926
; Sequence 15926, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15926
; LENGTH: 41435
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15926

Query Match          75.2%; Score 15.8; DB 3; Length 41435;
Best Local Similarity 89.5%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GCATGTATGGGTTATCTTC 20
      ||||| ||||| ||||| |||||
Db      40530 GCATATATGGGTTATTTTC 40548
```

```
RESULT 35
US-09-949-016-13845
; Sequence 13845, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13845
; LENGTH: 152132
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13845

Query Match      75.2%; Score 15.8; DB 3; Length 152132;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
Db      131531 TGAATGTATGGCTTATCTT 131549

RESULT 36
US-09-949-016-12371
; Sequence 12371, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12371
; LENGTH: 152145
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12371

Query Match      75.2%; Score 15.8; DB 3; Length 152145;
Best Local Similarity 89.5%; Pred. No. 4.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
Db      131531 TGAATGTATGGCTTATCTT 131549

RESULT 37
US-09-949-016-17371/c
; Sequence 17371, Application US/09949016
```

```
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17371
; LENGTH: 373182
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(373182)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17371

Query Match      75.2%; Score 15.8; DB 3; Length 373182;
Best Local Similarity 89.5%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
Db      66585 TGCATATATGAGTTATCTT 66567

RESULT 38
US-09-949-016-12062/c
; Sequence 12062, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12062
; LENGTH: 373694
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(373694)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12062

Query Match      75.2%; Score 15.8; DB 3; Length 373694;
Best Local Similarity 89.5%; Pred. No. 5.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTT 19
Db      66640 TGCATATATGAGTTATCTT 66622
```

```
RESULT 39
US-09-949-016-14577
; Sequence 14577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14577
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14577
Query Match 75.2%; Score 15.8; DB 3; Length 678533;
Best Local Similarity 89.5%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CATGTATGGGTATCTTCC 21
||| ||||| ||||| |||||
Db 590903 CATTATGGGTATTTTCC 590921

RESULT 40
US-09-949-016-14578
; Sequence 14578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14578
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14578
Query Match 75.2%; Score 15.8; DB 3; Length 678533;
Best Local Similarity 89.5%; Pred. No. 5.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CATGTATGGGTATCTTCC 21
||| ||||| ||||| |||||
Db 590903 CATTATGGGTATTTTCC 590921

RESULT 41
US-09-949-016-14033/c
; Sequence 14033, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14033
; LENGTH: 784019
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(784019)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14033
Query Match 75.2%; Score 15.8; DB 3; Length 784019;
Best Local Similarity 89.5%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTT 19
||||| ||||| ||||| |||||
Db 702418 TGCATGTATGTGTTTCTT 702400

RESULT 42
US-09-949-016-12777/c
; Sequence 12777, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12777
; LENGTH: 828152
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(828152)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12777
Query Match 75.2%; Score 15.8; DB 3; Length 828152;
Best Local Similarity 89.5%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

QY 1 TGCATGATGGGTATCTT 19  
Db 698551 TGCATGATGTTTCTT 698533

RESULT 43  
US-09-248-796A-5759/c  
; Sequence 5759, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstein et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.132  
; CURRENT APPLICATION NUMBER: US/09/248,796A  
; PRIOR FILING DATE: 1999-02-12  
; PRIOR APPLICATION NUMBER: US 60/074,725  
; PRIOR FILING DATE: 1998-02-13  
; PRIOR APPLICATION NUMBER: US 60/096,409  
; PRIOR FILING DATE: 1998-08-13  
; NUMBER OF SEQ ID NOS: 28208  
; SEQ ID NO 5759  
; LENGTH: 1965  
; TYPE: DNA  
; ORGANISM: Candida albicans  
US-09-248-796A-5759

Query Match 73.3%; Score 15.4; DB 3; Length 1965;  
Best Local Similarity 94.1%; Pred. No. 3.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTC 20  
Db 800 ATGTATGGTTATCTTC 784

RESULT 44  
US-09-949-016-15568/c  
; Sequence 15568, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15568  
; LENGTH: 91772  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(91772)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15568

Query Match 73.3%; Score 15.4; DB 3; Length 91772;  
Best Local Similarity 94.1%; Pred. No. 6.9e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 TGTATGGGTATCTTC 21  
Db 64103 TGTATGGTTATCTTC 64087

RESULT 45  
US-09-949-016-12393  
; Sequence 12393, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12393  
; LENGTH: 106199  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(106199)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12393

Query Match 73.3%; Score 15.4; DB 3; Length 106199;  
Best Local Similarity 94.1%; Pred. No. 7e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTC 20  
Db 24733 AIGTTGGGTATCTTC 24749

RESULT 46  
US-09-949-016-14370  
; Sequence 14370, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14370  
; LENGTH: 107980  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)...(107980)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14370

Query Match 73.3%; Score 15.4; DB 3; Length 107980;  
Best Local Similarity 94.1%; Pred. No. 7e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTC 20  
|||||  
Db 74514 AAGTTTGGGTATCTTC 74530

RESULT 47  
US-09-949-016-17009/c  
; Sequence 17009, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17009  
; LENGTH: 205163  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17009

Query Match 73.3%; Score 15.4; DB 3; Length 205163;  
Best Local Similarity 94.1%; Pred. No. 7.6e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATC 17  
|||||  
Db 76139 TGCATGTATGGGTAAAC 76123

RESULT 48  
US-08-286-819A-20/c  
; Sequence 20, Application US/08286819A  
; Patent No. 5871910  
; GENERAL INFORMATION:  
; APPLICANT: ARTHUR, MICHEL  
; APPLICANT: DUKTA-MALEN, SYLVIE  
; APPLICANT: MOLINAS, CATHERINE  
; APPLICANT: COURVALIN, PATRICE  
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
; EXPRESSION OF RESISTANCE TO GLYCOPEPTIDES, IN PARTICULAR  
; IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
; THESE POLYPEPTIDES AND USE FOR DIAGNOSIS  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/286,819A  
; FILING DATE: 05-AUG-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/174,682  
; FILING DATE: 28-DEC-1993

; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/917,146  
; FILING DATE: 10-AUG-1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/FR/91/00855  
; FILING DATE: 29-OCT-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FR 9013579  
; FILING DATE: 31-OCT-1990  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Oblon, No. 5871910man F.  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 413-3000  
; TELEFAX: (703) 413-2220  
; TELEX: 248855 OPAT UR  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 573 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..573  
US-08-286-819A-20

Query Match 72.4%; Score 15.2; DB 2; Length 573;  
Best Local Similarity 85.0%; Pred. No. 3.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
|||||  
Db 317 TGGCTGTATGGATTATCTTC 298

RESULT 49  
US-08-980-357-20/c  
; Sequence 20, Application US/08980357  
; Patent No. 6013508  
; GENERAL INFORMATION:  
; APPLICANT: ARTHUR, MICHEL  
; APPLICANT: DUKTA-MALEN, SYLVIE  
; APPLICANT: MOLINAS, CATHERINE  
; APPLICANT: COURVALIN, PATRICE  
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
; EXPRESSION OF RESISTANCE TO GLYCOPEPTIDES, IN PARTICULAR  
; IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
; THESE POLYPEPTIDES AND USE FOR DIAGNOSIS  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/980,357  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: US 08/286,819
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; APPLICATION DATA: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; APPLICATION NUMBER: FR 9013579
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NO. 6013508man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..573
;
US-08-980-357-20

Query Match 72.4%; Score 15.2; DB 3; Length 573;
Best Local Similarity 85.0%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTC 20
DB 317 TGGCTGTATGGATTATCTTC 298

RESULT 50
US-09-357-375-20/c
; Sequence 20, Application US/09357375
; Patent No. 6916906
; GENERAL INFORMATION:
; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE
; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR
; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/357,375
; FILING DATE:
; CLASSIFICATION:
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,819
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; APPLICATION DATA: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; APPLICATION NUMBER: FR 9013579
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NO. 6916906man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 573 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..573
;
US-09-357-375-20

Query Match 72.4%; Score 15.2; DB 3; Length 573;
Best Local Similarity 85.0%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTC 20
DB 317 TGGCTGTATGGATTATCTTC 298

RESULT 51
US-09-270-767-28376
; Sequence 28376, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28376
; LENGTH: 885
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
;
US-09-270-767-28376

Query Match 72.4%; Score 15.2; DB 3; Length 885;
Best Local Similarity 85.0%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTC 20
DB 698 TGGATATATCGGTTATCTTC 717

RESULT 52
US-09-270-767-12584
; Sequence 12584, Application US/09270767
```



TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7225 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-980-357-15

Query Match 72.4%; Score 15.2; DB 3; Length 7225;  
Best Local Similarity 85.0%; Pred. No. 5.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTATCTTC 20  
|| ||||| |||||  
Db 1048 TGGCTGATGGATTATCTTC 1029

RESULT 55  
US-09-357-375-15/c

Sequence 15, Application US/09357375  
Patent No. 6916906

GENERAL INFORMATION:

APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKTA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE

TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR  
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/357,375

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/286,819

FILING DATE: 05-AUG-1994

APPLICATION NUMBER: US 08/174,682

FILING DATE: 28-DEC-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917,146

FILING DATE: 10-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR/91/00855

FILING DATE: 29-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9013579

FILING DATE: 31-OCT-1990

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 6916906man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 660-060-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7225 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-357-375-15

Query Match 72.4%; Score 15.2; DB 3; Length 7225;  
Best Local Similarity 85.0%; Pred. No. 5.9e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTATCTTC 20  
|| ||||| |||||  
Db 1048 TGGCTGATGGATTATCTTC 1029

RESULT 56

US-08-286-819A-16/c

Sequence 16, Application US/08286819A

Patent No. 5871910

GENERAL INFORMATION:

APPLICANT: ARTHUR, MICHEL

APPLICANT: DUKTA-MALEN, SYLVIE

APPLICANT: MOLINAS, CATHERINE

APPLICANT: COURVALIN, PATRICE

TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE

EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR

TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR

TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESSEE: P.C.

STREET: 1755 S. Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/286,819A

FILING DATE: 05-AUG-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/174,682

FILING DATE: 28-DEC-1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917,146

FILING DATE: 10-AUG-1992

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR/91/00855

FILING DATE: 29-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9013579

FILING DATE: 31-OCT-1990

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 5871910man F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 660-060-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 413-3000

TELEFAX: (703) 413-2220

TELEX: 248855 OPAT UR



;  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10851 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-286-819A-16

Query Match 72.4%; Score 15.2; DB 2; Length 10851;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
DB 3503 TGGCTGTATGGATTATCTTC 3484

RESULT 57  
US-08-980-357-16/c

; Sequence 16, Application US/08980357  
; Patent No. 6013508

; GENERAL INFORMATION:

; APPLICANT: ARTHUR, MICHEL

; APPLICANT: DUKTA-MALEN, SYLVIE

; APPLICANT: MOLINAS, CATHERINE

; APPLICANT: COURVALIN, PATRICE

; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE

; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR

; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR

; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

; ADDRESSEE: P.C.

; STREET: 1755 S. Jefferson Davis Highway, Suite 400

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/980,357

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/286,819

; FILING DATE: 05-AUG-1994

; APPLICATION NUMBER: US 08/174,682

; FILING DATE: 28-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/917,146

; FILING DATE: 10-AUG-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/FR/91/00855

; FILING DATE: 29-OCT-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 9013579

; FILING DATE: 31-OCT-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Oblon, No. 6013508man F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 660-060-0 PCT

; TELEPHONE: (703) 413-3000

; TELEFAX: (703) 413-2220

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 10851 base pairs

; TYPE: nucleic acid

;  
; LENGTH: 10851 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-980-357-16

Query Match 72.4%; Score 15.2; DB 3; Length 10851;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
DB 3503 TGGCTGTATGGATTATCTTC 3484

RESULT 58

US-09-357-375-16/c

; Sequence 16, Application US/09357375

; Patent No. 6916906

; GENERAL INFORMATION:

; APPLICANT: ARTHUR, MICHEL

; APPLICANT: DUKTA-MALEN, SYLVIE

; APPLICANT: MOLINAS, CATHERINE

; APPLICANT: COURVALIN, PATRICE

; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE

; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPETIDES, IN PARTICULAR

; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR

; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS

; NUMBER OF SEQUENCES: 54

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

; ADDRESSEE: P.C.

; STREET: 1755 S. Jefferson Davis Highway, Suite 400

; CITY: Arlington

; STATE: Virginia

; COUNTRY: U.S.A.

; ZIP: 22202

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/357,375

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/286,819

; FILING DATE: 05-AUG-1994

; APPLICATION NUMBER: US 08/174,682

; FILING DATE: 28-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/917,146

; FILING DATE: 10-AUG-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/FR/91/00855

; FILING DATE: 29-OCT-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 9013579

; FILING DATE: 31-OCT-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Oblon, No. 6916906man F.

; REGISTRATION NUMBER: 24,618

; REFERENCE/DOCKET NUMBER: 660-060-0 PCT

; TELEPHONE: (703) 413-3000

; TELEFAX: (703) 413-2220

; INFORMATION FOR SEQ ID NO: 16:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 10851 base pairs

; TYPE: nucleic acid

STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-09-357-375-16

Query Match 72.4%; Score 15.2; DB 3; Length 10851;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
||| ||||| ||||| |||||  
Db 3503 TGGCTGATGGATTATCTTC 3484

RESULT 59  
US-09-949-016-11748  
; Sequence 11748, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11748  
; LENGTH: 29121  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11748

Query Match 72.4%; Score 15.2; DB 3; Length 29121;  
Best Local Similarity 85.0%; Pred. No. 7.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21  
||| ||||| ||||| |||||  
Db 13951 GCATGTATTGTTATGTTCC 13970

RESULT 60  
US-09-949-016-13591  
; Sequence 13591, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13591  
; LENGTH: 29122  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13591

Query Match 72.4%; Score 15.2; DB 3; Length 29122;  
Best Local Similarity 85.0%; Pred. No. 7.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21  
||| ||||| ||||| |||||  
Db 13951 GCATGTATTGTTATGTTCC 13970

Search completed: January 28, 2006, 01:18:37  
Job time : 66.9888 secs

This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (C) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:15:53 ; Search time 231.809 Seconds  
(without alignments)  
82.419 Million cell updates/sec

Title: US-10-716-005-1

Perfect score: 23

Sequence: 1 tgaagaggcagtagaagcttag 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6059551 seqs, 41533918 residues

Total number of hits satisfying chosen parameters: 12119102

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA New.\*  
1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US12\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US13\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US14\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US15\_NEW\_PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US16\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.8	81.7	201	7	US-10-995-561-48428
2	18.8	81.7	81210	7	US-10-995-561-13295
3	18.8	81.7	171732	8	US-11-121-086-98
4	17.2	74.8	24	7	US-10-310-914A-636042
5	17.2	74.8	886	5	US-09-978-360A-196
6	17.2	74.8	17517	8	US-11-136-527-3650
7	16.8	73.0	769	7	US-10-750-185-43151
8	16.8	73.0	769	7	US-10-750-623-43151
9	16.8	73.0	150450	8	US-11-112-908-54
10	16.8	73.0	191343	8	US-11-112-908-53
11	16.6	72.2	201	8	US-11-124-368A-20510
12	16.6	72.2	201	8	US-11-124-368A-20540
13	16.6	72.2	600	7	US-10-750-185-2581
14	16.6	72.2	600	7	US-10-750-185-2668
15	16.6	72.2	600	7	US-10-750-623-2581
16	16.6	72.2	600	7	US-10-750-623-2668
17	16.6	72.2	600	8	US-11-136-527-7331
18	16.6	72.2	933	7	US-10-750-185-40685
19	16.6	72.2	933	7	US-10-750-623-40685
20	16.6	72.2	1962	8	US-11-136-527-3235
21	16.6	72.2	2200	7	US-10-750-185-59759
22	16.6	72.2	2200	7	US-10-750-623-59759

23	16.6	72.2	3257	8	US-11-128-061-3404	Sequence 3404, Ap
24	16.6	72.2	3257	8	US-11-128-049-3404	Sequence 3404, Ap
25	16.6	72.2	41081	8	US-11-124-368A-2931	Sequence 2931, Ap
26	16.6	72.2	165627	7	US-11-121-086-89	Sequence 89, Appl
27	16.4	71.3	201	7	US-10-995-561-59465	Sequence 59465, A
28	16.4	71.3	600	8	US-11-136-527-6022	Sequence 6022, Ap
29	16.4	71.3	2391	8	US-11-136-527-1926	Sequence 1926, Ap
30	16.4	71.3	83528	7	US-10-995-561-13343	Sequence 13343, A
31	16.4	71.3	162085	8	US-11-121-086-7	Sequence 7, Appl
32	16.2	70.4	21	7	US-10-310-914A-635991	Sequence 635991, A
33	16.2	70.4	25	8	US-11-121-849-93224	Sequence 93224, A
34	16.2	70.4	25	8	US-11-121-849-94187	Sequence 94187, A
35	16.2	70.4	201	7	US-10-995-561-27893	Sequence 27893, A
36	16.2	70.4	201	7	US-10-995-561-72197	Sequence 72197, A
37	16.2	70.4	482	8	US-11-128-061-2983	Sequence 2983, Ap
38	16.2	70.4	482	8	US-11-128-061-6625	Sequence 6625, Ap
39	16.2	70.4	482	8	US-11-128-049-2983	Sequence 2983, Ap
40	16.2	70.4	482	8	US-11-128-049-6625	Sequence 6625, Ap
41	16.2	70.4	984	7	US-10-750-185-29356	Sequence 29356, A
42	16.2	70.4	984	7	US-10-750-623-29356	Sequence 29356, A
43	16.2	70.4	7320	7	US-10-486-968-1	Sequence 1, Appl
44	16.2	70.4	47622	7	US-10-995-561-13431	Sequence 13431, A
45	16.2	70.4	53328	7	US-10-995-561-13243	Sequence 13243, A
46	16	69.6	22	7	US-10-310-914A-1196435	Sequence 1196435, A
47	16	69.6	24	7	US-10-310-914A-1196435	Sequence 1196435, A
48	15.8	68.7	28	7	US-10-310-914A-406624	Sequence 406624, A
49	15.8	68.7	1041	7	US-10-750-185-63692	Sequence 63692, A
50	15.8	68.7	1041	7	US-10-750-623-63692	Sequence 63692, A
51	15.8	68.7	1066	7	US-10-750-185-26352	Sequence 26352, A
52	15.8	68.7	1066	7	US-10-750-623-26352	Sequence 26352, A
53	15.8	68.7	73404	8	US-11-124-368A-2914	Sequence 2914, Ap
54	15.8	68.7	190376	7	US-10-661-966-1	Sequence 1, Appl
55	15.8	68.7	191350	7	US-10-857-780-4	Sequence 4, Appl
56	15.8	68.7	1691140	8	US-11-091-018-1	Sequence 1, Appl
57	15.6	67.8	22	7	US-10-310-914A-156548	Sequence 156548, A
58	15.6	67.8	25	8	US-11-136-527-79009	Sequence 79009, A
59	15.6	67.8	25	8	US-11-136-527-79014	Sequence 79014, A
60	15.6	67.8	25	8	US-11-136-527-278745	Sequence 278745, A

ALIGNMENTS

RESULT 1

US-10-995-561-48428/c  
; Sequence 48428, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF DETECTION AND USES THEREOF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CLO01559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48428  
; LENGTH: 201  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-995-561-48428

Query Match 81.7%; Score 18.8; DB 7; Length 201;  
Best Local Similarity 90.9%; Pred. No. 10;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAAGGCAGTAGAAGCTTAG 23

Db 25 GAGAAGGAGTAGAAGCCTTAG 4

RESULT 2

```
US-10-995-561-13295/c
; Sequence 13295, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13295
; LENGTH: 81210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13295
Query Match      81.7%; Score 18.8; DB 7; Length 81210;
Best Local Similarity 90.9%; Pred. No. 29;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAGAGGCGAGTAGAAAGCTTAG 23
Db 525 GAGAGGCGAGTAGAAAGCTTAG 504
RESULT 3
US-11-121-086-98/c
; Sequence 98, Application US/11121086
; Publication No. US220050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 98
; LENGTH: 171732
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-98
Query Match      81.7%; Score 18.8; DB 8; Length 171732;
Best Local Similarity 90.9%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 GAGAGGCGAGTAGAAAGCTTAG 23
Db 131125 GAGAGGCGAGTAGAAAGCTTAG 131104
RESULT 4
US-10-310-914A-636042
; Sequence 636042, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiller, Kvazat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 636042
```

```
; LENGTH: 24
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-636042
Query Match      74.8%; Score 17.2; DB 7; Length 24;
Best Local Similarity 81.8%; Pred. No. 40;
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 2 GAGAAGGCGAGTAGAAAGCTTAG 23
Db 2 GAGAAGGCGAGTAGAAAGGAG 23
RESULT 5
US-09-978-360A-196/c
; Sequence 196, Application US/09978360A
; Publication No. US20060009633A9
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste Dumas Milne
; APPLICANT: Ducier, Aymeric
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; APPLICANT: Clusel, Catherine
; TITLE OF INVENTION: Complementary DNA's Encoding Proteins with Signal Peptides
; FILE REFERENCE: 56.US4.CIP
; CURRENT APPLICATION NUMBER: US/09/978,360A
; CURRENT FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/066,677
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: US 60/069,957
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: US 60/074,121
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: US 60/081,563
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: US 60/096,116
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: US 60/099,273
; PRIOR FILING DATE: -09-04
; PRIOR APPLICATION NUMBER: US 09/191,997
; PRIOR FILING DATE: 1998-11-13
; PRIOR APPLICATION NUMBER: US 09/215,435
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: PCT/IB98/02122
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: US 09/247,155
; PRIOR FILING DATE: 1999-02-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 810
; SOFTWARE: Patent.pm
; SEQ ID NO 196
; LENGTH: 886
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31..381
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 31..90
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.4
; OTHER INFORMATION: seq AFVIACVLSLIST/IY
; FEATURE:
; NAME/KEY: polyA site
; LOCATION: 875..886
US-09-978-360A-196
Query Match      74.8%; Score 17.2; DB 5; Length 886;
Best Local Similarity 86.4%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 2 GAGAGGCGAGTGAAGCTTAG 23  
Db 473 GAGAGGCGAGTGAAGAGAG 452

RESULT 6  
US-11-136-527-3650/c  
; Sequence 3650, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3650  
; LENGTH: 17517  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-3650

Query Match 74.8%; Score 17.2; DB 8; Length 17517;  
Best Local Similarity 86.4%; Pred. No. 1.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTGAAGCTTAG 23  
Db 13480 GAGAGGCGAGGCAAGCTGAG 13459

RESULT 7  
US-10-750-185-43151  
; Sequence 43151, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 43151  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Bovine 19866881328491  
US-10-750-185-43151

Query Match 73.0%; Score 15.8; DB 7; Length 769;  
Best Local Similarity 90.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTGAAGCTT 21  
Db 660 GAGATGCGAGCAAGCTT 679

RESULT 8  
US-10-750-623-43151  
; Sequence 43151, Application US/10750623

; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 43151  
; LENGTH: 769  
; TYPE: DNA  
; ORGANISM: Bovine 19866881328491  
US-10-750-623-43151

Query Match 73.0%; Score 16.8; DB 7; Length 769;  
Best Local Similarity 90.0%; Pred. No. 1.1e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTGAAGCTT 21  
Db 660 GAGATGCGAGCAAGCTT 679

RESULT 9  
US-11-112-908-54  
; Sequence 54, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Cole  
; APPLICANT: Davis, Lisa M.  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 54  
; LENGTH: 150450  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-54

Query Match 73.0%; Score 16.8; DB 8; Length 150450;  
Best Local Similarity 90.0%; Pred. No. 2.7e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAGGCGAGTGAAGCTTAG 23  
Db 145183 GAGACAGGAGAAAGCTTAG 145202

RESULT 10  
US-11-112-908-53  
; Sequence 53, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:

```
; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 53
; LENGTH: 191343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-53

Query Match      73.0%; Score 16.8; DB 8; Length 191343;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAAGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 64425 GAAGACAGGAGAAAGCTTAG 64444

RESULT 11
US-11-124-368A-20510
; Sequence 20510, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20510
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-20510

Query Match      72.2%; Score 16.6; DB 8; Length 201;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 135 TGGGAAGGAAGGAGAAAGCTTAG 157

RESULT 12
US-11-124-368A-20540
; Sequence 20540, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20510
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-20510

Query Match      72.2%; Score 16.6; DB 8; Length 201;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 135 TGGGAAGGAAGGAGAAAGCTTAG 157

; APPLICANT: Harris, Cole
; APPLICANT: Davis, Lisa M.
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 53
; LENGTH: 191343
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-53

Query Match      73.0%; Score 16.8; DB 8; Length 191343;
Best Local Similarity 90.0%; Pred. No. 2.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAAGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 64425 GAAGACAGGAGAAAGCTTAG 64444

RESULT 11
US-11-124-368A-20510
; Sequence 20510, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20510
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-20510

Query Match      72.2%; Score 16.6; DB 8; Length 201;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 135 TGGGAAGGAAGGAGAAAGCTTAG 157

RESULT 12
US-11-124-368A-20540
; Sequence 20540, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20540
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-20540

Query Match      72.2%; Score 16.6; DB 8; Length 201;
Best Local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 105 TGGGAAGGAAGGAGAAAGCTTAG 127

RESULT 13
US-10-750-185-2581
; Sequence 2581, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2581
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine
MMBT13689
US-10-750-185-2581

Query Match      72.2%; Score 16.6; DB 7; Length 600;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCAGTGTAGAAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 425 TGTGAAGGCAGGAGAAAAGCTGAG 447

RESULT 14
US-10-750-185-2668
; Sequence 2668, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
```

; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2668  
; TYPE: DNA  
; ORGANISM: Bovine MMBT08581  
US-10-750-185-2668

Query Match 72.2%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 443 TGTGAAGGCAGGAGAAAGCTGAG 465

RESULT 15  
US-10-750-623-2581  
; Sequence 2581, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2581  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT13689  
US-10-750-623-2581

Query Match 72.2%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 425 TGTGAAGGCAGGAGAAAGCTGAG 447

RESULT 16  
US-10-750-623-2668  
; Sequence 2668, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623

; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2668  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Bovine MMBT08581  
US-10-750-623-2668

Query Match 72.2%; Score 16.6; DB 7; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 443 TGTGAAGGCAGGAGAAAGCTGAG 465

RESULT 17  
US-11-136-527-7331  
; Sequence 7331, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7331  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
US-11-136-527-7331

Query Match 72.2%; Score 16.6; DB 8; Length 600;  
Best Local Similarity 82.6%; Pred. No. 1.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 484 TGAAGAGGTAGCAGCAAGCTGAG 506

RESULT 18  
US-10-750-185-40685/c  
; Sequence 40685, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 40685  
; LENGTH: 933  
; TYPE: DNA

```

; ORGANISM: Bovine 19866880900388
US-10-750-185-40685

Query Match      72.2%; Score 16.6; DB 7; Length 933;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
DB 867 TGACACGGAAGAAGAAAGCTTAG 845

RESULT 19
US-10-750-623-40685/c
; Sequence 40685, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 40685
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Bovine 19866880900388
US-10-750-623-40685

Query Match      72.2%; Score 16.6; DB 7; Length 933;
Best Local Similarity 82.6%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
DB 867 TGACACGGAAGAAGAAAGCTTAG 845

RESULT 20
US-11-136-527-3235
; Sequence 3235, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3235
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-3235

Query Match      72.2%; Score 16.6; DB 8; Length 1962;
Best Local Similarity 82.6%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23

```

```

DB 1846 TGAGAAGGTAGCAGCAAGCTGAG 1868

RESULT 21
US-10-750-185-59759
; Sequence 59759, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59759
; LENGTH: 2200
; TYPE: DNA
; ORGANISM: Bovine 19866880587237
US-10-750-185-59759

Query Match      72.2%; Score 16.6; DB 7; Length 2200;
Best Local Similarity 82.6%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
DB 1837 TGATGAGGCAGTAGAAGGCTAG 1859

RESULT 22
US-10-750-623-59759
; Sequence 59759, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59759
; LENGTH: 2200
; TYPE: DNA
; ORGANISM: Bovine 19866880587237
US-10-750-623-59759

Query Match      72.2%; Score 16.6; DB 7; Length 2200;
Best Local Similarity 82.6%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
DB 1837 TGATGAGGCAGTAGAAGGCTAG 1859

```



```
RESULT 23
US-11-128-061-3404
; Sequence 3404, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; PRIOR FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3404
; LENGTH: 3257
; TYPE: DNA
; ORGANISM: Cricetus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1066)..(1086)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2121)..(2140)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-3404

Query Match          72.2%; Score 16.6; DB 8; Length 3257;
Best Local Similarity 82.6%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  TCAGAAGCCGAGTAAAGCTTAG 23
Db      1585 TCAGAAGCCGAGTAAAGCTTAG 1607

RESULT 24
US-11-128-049-3404
; Sequence 3404, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3404
; LENGTH: 3257
; TYPE: DNA
; ORGANISM: Cricetus griseus
; FEATURE:
```

```
; NAME/KEY: misc feature
; LOCATION: (1066)..(1086)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (2121)..(2140)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-3404

Query Match          72.2%; Score 16.6; DB 8; Length 3257;
Best Local Similarity 82.6%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  TCAGAAGCCGAGTAAAGCTTAG 23
Db      1585 TCAGAAGCCGAGTAAAGCTTAG 1607

RESULT 25
US-11-124-368A-2931
; Sequence 2931, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2931
; LENGTH: 41081
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-2931

Query Match          72.2%; Score 16.6; DB 8; Length 41081;
Best Local Similarity 82.6%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  TCAGAAGCCGAGTAAAGCTTAG 23
Db      27929 TCGGAAGGAAGGAGAAAGCTTAG 27951

RESULT 26
US-11-121-086-89
; Sequence 89, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 89
; LENGTH: 165627
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-89
```

```
Query Match          72.2%; Score 16.6; DB 8; Length 165627;
Best Local Similarity 82.6%; Pred. No. 3.4e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
   ||||| ||| ||| |||
Db 101723 TAAGAAGGCAGCAGACATCTTAG 101745

RESULT 27
US-10-995-561-59465/c
; Sequence 59465, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 59465
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-59465

Query Match          71.3%; Score 16.4; DB 7; Length 201;
Best Local Similarity 85.0%; Pred. No. 1.4e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTA 22
   ||||| ||| ||| |||
Db 119 AGCAGGCAGTAGAAGATYA 100

RESULT 28
US-11-136-527-6022/c
; Sequence 6022, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6022
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6022

Query Match          71.3%; Score 16.4; DB 8; Length 600;
Best Local Similarity 77.3%; Pred. No. 1.6e+02;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
   ||||| ||| ||| |||
Db 145 GAGAAGGCCCAASAARGCTTAG 124

RESULT 29
US-11-136-527-1926/c
; Sequence 1926, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1926
; LENGTH: 2391
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-1926

Query Match          71.3%; Score 16.4; DB 8; Length 2391;
Best Local Similarity 77.3%; Pred. No. 2.1e+02;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
   ||||| ||| ||| |||
Db 1936 GAGAAGGCCCAASAARGCTTAG 1915

RESULT 30
US-10-995-561-13343/c
; Sequence 13343, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13343
; LENGTH: 83528
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13343

Query Match          71.3%; Score 16.4; DB 7; Length 83528;
Best Local Similarity 85.0%; Pred. No. 3.8e+02;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTA 22
   ||||| ||| ||| |||
Db 16778 AGCAGGCAGTAGAAGATYA 16759

RESULT 31
US-11-121-086-7/c
; Sequence 7, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138-6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7
; LENGTH: 162085
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
US-11-121-086-7
Query Match          71.3%; Score 16.4; DB 8; Length 162085;
Best Local Similarity 94.4%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      4 GAAGGCAGTAGAAGCTT 21
      |||||
Db      3544 GAAGGCAGTAGAAGCTT 3527

RESULT 32
US-10-310-914A-635991
; Sequence 635991, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/10/310,914A
; PRIOR FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 635991
; LENGTH: 21
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-635991

Query Match          70.4%; Score 16.2; DB 7; Length 21;
Best Local Similarity 81.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAGCTTAG 23
      |||||
Db      1 AGAAGGCAGTAGAAGCTTAG 21

RESULT 33
US-11-121-849-93224
; Sequence 93224, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; PRIOR FILING DATE: 2005-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 93224
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-93224

Query Match          70.4%; Score 16.2; DB 8; Length 25;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAAGGCAGTAGAAGCTT 21
      |||||
Db      3 TGAGAAGGCAGTAGAAGCTT 23

RESULT 34
US-11-121-849-94187
; Sequence 94187, Application US/11121849
```

```
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded ;
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; PRIOR FILING DATE: 2005-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 94187
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-94187

Query Match          70.4%; Score 16.2; DB 8; Length 25;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGAGAAGGCAGTAGAAGCTT 21
      |||||
Db      3 TGAGAAGGCAGTAGAAGCTT 23

RESULT 35
US-10-995-561-27893
; Sequence 27893, Application US/109955561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; PRIOR FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27893
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-27893

Query Match          70.4%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAGCTTAG 23
      |||||
Db      105 AGCAGGCTGTAGAGAGCTTAG 125

RESULT 36
US-10-995-561-72197
; Sequence 72197, Application US/109955561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; PRIOR FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 72197
; LENGTH: 201
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-995-561-72197

Query Match      70.4%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTAG 23
DB 105 AGCAGGCTGTAGAGAGCTTAG 125

RESULT 37
US-11-128-061-2983
; Sequence 2983, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2983
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; NAME/KEY: misc feature
; LOCATION: (418)..(446)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-2983

Query Match      70.4%; Score 16.2; DB 8; Length 482;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
DB 5 TGAGAAGGCAGTGCAAAAGCT 25

RESULT 38
US-11-128-061-6625
; Sequence 6625, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
```

```
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6625
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418)..(446)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-6625

Query Match      70.4%; Score 16.2; DB 8; Length 482;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
DB 5 TGAGAAGGCAGTGCAAAAGCT 25

RESULT 39
US-11-128-049-2983
; Sequence 2983, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; TITLE OF INVENTION: MAKING AND USING SAME
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 2983
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418)..(446)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-2983

Query Match      70.4%; Score 16.2; DB 8; Length 482;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
DB 5 TGAGAAGGCAGTGCAAAAGCT 25

RESULT 40
US-11-128-049-6625
; Sequence 6625, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
```

```
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 6625
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418)..(446)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-6625

Query Match          70.4%; Score 16.2; DB 8; Length 482;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTT 21
   ||||| ||||| ||||| |||||
DB 5 TGAGAGGCGAGTGCAAGCCT 25

RESULT 41
US-10-750-185-29356/c
; Sequence 29356, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29356
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Bovine 19866881161138
US-10-750-185-29356

Query Match          70.4%; Score 16.2; DB 7; Length 984;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGAAGGCGAGTAGAAGCTTAG 23
   ||||| ||||| ||||| |||||
DB 918 AGAAGGCGAGTAAAAAGTTAG 898

RESULT 42
US-10-750-623-29356/c
; Sequence 29356, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29356
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Bovine 19866881161138
US-10-750-623-29356

Query Match          70.4%; Score 16.2; DB 7; Length 984;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGAAGGCGAGTAGAAGCTTAG 23
   ||||| ||||| ||||| |||||
DB 918 AGAAGGCGAGTAAAAAGTTAG 898

RESULT 43
US-10-486-968-1
; Sequence 1, Application US/10486968
; Publication No. US20050250205A1
; GENERAL INFORMATION:
; APPLICANT: Genoplate-Valor S.A.S
; TITLE OF INVENTION: Use of associations between at least one nucleic sequence
; TITLE OF INVENTION: polymorphism of the SH2 gene and at least one seed
; FILE REFERENCE: SH2
; CURRENT APPLICATION NUMBER: US/10/486,968
; CURRENT FILING DATE: 2004-02-17
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 7320
; TYPE: DNA
; ORGANISM: Zea mays
US-10-486-968-1

Query Match          70.4%; Score 16.2; DB 7; Length 7320;
Best Local Similarity 85.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AGAAGGCGAGTAGAAGCTTAG 23
   ||||| ||||| ||||| |||||
DB 1647 AGAAGCGAGGAGAAAGCTTTG 1667

RESULT 44
US-10-995-561-13431
; Sequence 13431, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13431
; LENGTH: 47622
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-995-561-13431

Query Match      70.4%; Score 16.2; DB 7; Length 47622;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTTAG 23
Db      38922 AGCAGGCTGTAGAGAGCTTAG 38942

RESULT 45
US-10-995-561-13243
; Sequence 13243, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13243
; LENGTH: 53328
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13243

Query Match      70.4%; Score 16.2; DB 7; Length 53328;
Best Local Similarity 85.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTTAG 23
Db      614 AGCAGGCTGTAGAGAGCTTAG 634

RESULT 46
US-10-914A-1196483/c
; Sequence 1196483, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiller, Kvuzat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1196483
; LENGTH: 22
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-1196483

Query Match      69.6%; Score 16; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAG 18
Db      22 AGAAGGCAGTAGAAAG 7

RESULT 47
US-10-310-914A-1196435/c
; Sequence 1196435, Application US/10310914A
```

```
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiller, Kvuzat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1196435
; LENGTH: 24
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-1196435

Query Match      69.6%; Score 16; DB 7; Length 24;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAG 18
Db      18 AGAAGGCAGTAGAAAG 3

RESULT 48
US-10-310-914A-406624
; Sequence 406624, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiller, Kvuzat
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 406624
; LENGTH: 28
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-406624

Query Match      68.7%; Score 15.8; DB 7; Length 28;
Best Local Similarity 73.7%; Pred. No. 1.8e+02;
Matches 14; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGAAGGCAGTAGAAAGCTT 21
Db      6 AGAAGGCAGUAUAAACCUU 24

RESULT 49
US-10-750-185-63692
; Sequence 63692, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
```



```

Query Match      68.7%; Score 15.8; DB 8; Length 73404;
Best Local Similarity 89.5%; Pred. No. 6.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCGAGTGAAGC 19
    ||||| ||||| ||||| |||||
Db 9784 TGAGAGGCGAGGAGAAAGC 9802

RESULT 54
US-10-661-966-1
; Sequence 1, Application US/10661966
; Publication No. US2005027118A1
; GENERAL INFORMATION:
; APPLICANT: Roth, Richard B.
; APPLICANT: Nelson, Matthew Roberts
; APPLICANT: Braun, Andreas
; TITLE OF INVENTION: METHODS FOR IDENTIFYING SUBJECTS AT RISK
; TITLE OF INVENTION: OF MELANOMA AND TREATMENTS THEREOF
; FILE REFERENCE: 524592003800
; CURRENT APPLICATION NUMBER: US/10/661,966
; CURRENT FILING DATE: 2003-09-11
; PRIOR APPLICATION NUMBER: 60/410,595
; PRIOR FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 60/422,344
; PRIOR FILING DATE: 2002-10-29
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 190276
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-661-966-1

Query Match      68.7%; Score 15.8; DB 7; Length 190276;
Best Local Similarity 89.5%; Pred. No. 7.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AGAAGGCGAGTGAAGCCTT 21
    ||||| ||||| ||||| |||||
Db 113905 AGAAGGCGAGGAGAAAGCCTT 113923

RESULT 55
US-10-857-780-4/c
; Sequence 4, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOVAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 191350
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```

```

; NAME/KEY: misc feature
; LOCATION: (16914)..(16914)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (19266)..(19266)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (26334)..(26334)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (34472)..(34472)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (38627)..(38628)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40555)..(40555)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (57355)..(57355)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (61207)..(61207)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (64980)..(64980)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (64987)..(64987)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (81452)..(81452)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (95893)..(95893)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (103359)..(103359)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (117565)..(117565)
; OTHER INFORMATION: n is a, c, g, or t
US-10-857-780-4

Query Match      68.7%; Score 15.8; DB 7; Length 191350;
Best Local Similarity 81.0%; Pred. No. 7.7e+02;
Matches 17; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCGAGTGAAGCCTT 21
    ||||| ||||| ||||| |||||
Db 17982 TGAGAGGCGCGAGAAAGCCTT 17962

RESULT 56
US-11-091-018-1
; Sequence 1, Application US/11091018
; Publication No. US20050287551A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
; APPLICANT: Thorleifsson, Gudmar

```



```
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: SUSCEPTIBILITY GENE FOR HUMAN STROKE;
; FILE REFERENCE: 2345.2010-016
; CURRENT APPLICATION NUMBER: US/11/091,018
; CURRENT FILING DATE: 2005-03-25
; PRIOR APPLICATION NUMBER: PCT/US03/29906
; PRIOR FILING DATE: 2003-09-25
; PRIOR APPLICATION NUMBER: 10/255,120
; PRIOR FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 10/419,723
; PRIOR FILING DATE: 2003-04-18
; PRIOR APPLICATION NUMBER: 10/650,120
; PRIOR FILING DATE: 2003-08-27
; PRIOR APPLICATION NUMBER: 10/067,514
; PRIOR FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: 09/811,352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691140
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(1691140)
; OTHER INFORMATION: n=A,T,C or G
US-11-091-018-1

Query Match      68.7%; Score 15.8; DB 8; Length 1691140;
Best Local Similarity 89.5%; Pred. No. 4.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      5 AAGCAGTGAAGAGCTTAG 23
Db      798654 AAGCAGTGAAGAGCTTAG 798672

RESULT 57
US-10-310-914A-156548
; Sequence 156548, Application US/10310914A
; Publication No. US20060003322A1
; GENERAL INFORMATION:
; APPLICANT: Bentwich, Isaac
; APPLICANT: Shiler, Kvuza
; TITLE OF INVENTION: Bioinformatically detectable group of novel regulatory genes and
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: 06087.0200.CPUS01
; CURRENT APPLICATION NUMBER: US/10/310,914A
; CURRENT FILING DATE: 2002-12-06
; NUMBER OF SEQ ID NOS: 1388402
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 156548
; LENGTH: 22
; TYPE: RNA
; ORGANISM: Human
US-10-310-914A-156548

Query Match      67.8%; Score 15.6; DB 7; Length 22;
Best Local Similarity 68.2%; Pred. No. 2.2e+02;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGAGAAGGCAGTGAAGAGCTTA 22
Db      1 UGAGAAGGCAGTGAAGAGCTTA 22

RESULT 58
US-11-136-527-79009
; Sequence 79009, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79009
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-79009
```

Query Match 67.8%; Score 15.6; DB 8; Length 25;  
Best Local Similarity 81.8%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTGAAGAGCTTAG 23  
Db 1 GATCAGGCAGTGAAGAGCTTAG 22

```
RESULT 59
US-11-136-527-79014
; Sequence 79014, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 79014
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-79014
```

Query Match 67.8%; Score 15.6; DB 8; Length 25;  
Best Local Similarity 81.8%; Pred. No. 2.2e+02;  
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTGAAGAGCTTAG 23  
Db 2 GATCAGGCAGTGAAGAGCTTAG 23

```
RESULT 60
US-11-136-527-278745
; Sequence 278745, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 278745
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-278745

Query Match      67.8%; Score 15.6; DB 8; Length 25;
Best Local Similarity 81.8%; Pred. NO. 2.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAGGCTTAG 23
      || ||||| ||||| |||||
Db       4 GATCAGGCGAGAGAGGCTTAG 25

Search completed: January 28, 2006, 02:00:36
Job time : 240.059 secs
```

This Page Blank (uspto)

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:01:27 ; Search time 399.011 Seconds  
(without alignments)  
476.668 Million cell updates/sec

Title: US-10-716-005-1

Perfect score: 23

Sequence: 1 tgaagaggcagtagaagcttag 23

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

- Published Applications NA Main:\*
- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
  - 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
  - 3: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
  - 4: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
  - 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
  - 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
  - 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
  - 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
  - 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
  - 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23	100.0	23	9	US-10-716-005-1
2	19.8	86.1	1378	7	US-10-465-205-14
3	19.8	86.1	1379	7	US-10-465-205-7
4	19.8	86.1	1379	7	US-10-465-205-13
5	19.8	86.1	1379	7	US-10-465-205-16
6	19.8	86.1	1384	7	US-10-465-205-8
7	19.8	86.1	1384	7	US-10-465-205-10
8	19.8	86.1	1385	7	US-10-465-205-9
9	19.8	86.1	1390	7	US-10-465-205-11
10	19.8	86.1	1393	7	US-10-465-205-15
11	19.8	86.1	1447	7	US-10-465-205-17
12	19.8	86.1	1734	7	US-10-282-122A-38442
13	19.8	86.1	1803	7	US-10-465-205-5
14	18.8	81.7	81210	7	US-10-741-601-5664
15	18.8	81.7	81210	8	US-10-741-600-17691
16	18.2	79.1	720	9	US-10-504-582-15
17	18.2	79.1	160820	8	US-10-473-939-5
18	17.8	77.4	522	4	US-09-925-065A-739139
19	17.8	77.4	525	4	US-09-925-065A-739199
20	17.8	77.4	545	4	US-09-925-065A-728631
21	17.8	77.4	624	4	US-09-925-065A-69405
22	17.8	77.4	863	4	US-09-925-065A-8982
23	17.8	77.4	1206	4	US-09-925-065A-709349

C	24	17.8	77.4	1206	4	US-09-925-065A-709350	Sequence 709350,
	25	17.8	77.4	1253	4	US-09-925-065A-62246	Sequence 62246, A
	26	17.8	77.4	1714	7	US-10-425-114-24004	Sequence 24004, A
C	27	17.8	77.4	1940	5	US-10-027-632-97987	Sequence 97987, A
C	28	17.8	77.4	1940	6	US-10-027-632-97987	Sequence 97987, A
	29	17.8	77.4	1991	8	US-10-425-115-112272	Sequence 112272, A
	30	17.8	77.4	2478	7	US-10-437-963-95439	Sequence 95439, A
C	31	17.8	77.4	2533	6	US-10-108-260A-1480	Sequence 1480, Ap
C	32	17.8	77.4	3279	5	US-10-027-632-114129	Sequence 114128,
C	33	17.8	77.4	3279	5	US-10-027-632-114129	Sequence 114129,
C	34	17.8	77.4	3279	6	US-10-027-632-114128	Sequence 114128,
C	35	17.8	77.4	3279	6	US-10-027-632-114128	Sequence 114128,
	36	17.8	77.4	6805	10	US-11-097-143-16906	Sequence 16906, A
C	37	17.8	77.4	17391	7	US-10-741-601-5754	Sequence 5754, Ap
	38	17.8	77.4	48265	7	US-10-741-601-5677	Sequence 5677, Ap
C	39	17.8	77.4	104399	7	US-10-322-281-123	Sequence 123, App
	40	17.4	75.7	21	7	US-10-465-205-1	Sequence 1, Appl
C	41	17.4	75.7	309	7	US-10-469-285-225	Sequence 225, App
C	42	17.4	75.7	322	3	US-09-917-800A-1065	Sequence 1065, Ap
C	43	17.4	75.7	569	4	US-09-925-065A-560212	Sequence 560212,
	44	17.4	75.7	2151	9	US-10-764-420-2470	Sequence 2470, Ap
	45	17.4	75.7	24789	10	US-11-097-143-40201	Sequence 40201, A
	46	17.2	74.8	300	7	US-10-242-535A-45247	Sequence 45247, A
	47	17.2	74.8	300	7	US-10-085-783A-45247	Sequence 45247, A
C	48	17.2	74.8	494	5	US-10-062-254-187	Sequence 187, App
	49	17.2	74.8	519	3	US-09-917-800A-436	Sequence 436, App
	50	17.2	74.8	546	5	US-10-027-632-237122	Sequence 237122,
	51	17.2	74.8	546	6	US-10-027-632-237122	Sequence 237122,
	52	17.2	74.8	568	5	US-10-027-632-49562	Sequence 49562, A
	53	17.2	74.8	568	5	US-10-027-632-49563	Sequence 49563, A
	54	17.2	74.8	568	5	US-10-027-632-49564	Sequence 49564, A
	55	17.2	74.8	568	6	US-10-027-632-49562	Sequence 49562, A
	56	17.2	74.8	568	6	US-10-027-632-49563	Sequence 49563, A
	57	17.2	74.8	568	6	US-10-027-632-49564	Sequence 49564, A
	58	17.2	74.8	583	4	US-09-925-065A-34	Sequence 34, Appl
	59	17.2	74.8	600	9	US-10-972-079-87189	Sequence 87189, A
	60	17.2	74.8	600	9	US-10-972-079-87190	Sequence 87190, A

ALIGNMENTS

RESULT 1  
US-10-716-005-1  
; Sequence 1, Application US/10716005  
; Publication No. US20050106578A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; APPLICANT: Cockerill III, Franklin R.  
; APPLICANT: Aichinger, Christian  
; APPLICANT: Reiser, Astrid  
; TITLE OF INVENTION: Detection of Group B Streptococcus  
; FILE REFERENCE: 07039/460001  
; CURRENT APPLICATION NUMBER: US/10/716,005  
; CURRENT FILING DATE: 2003-11-18  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 23  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide  
US-10-716-005-1

Query Match 100.0%; Score 23; DB 9; Length 23;  
Best Local Similarity 100.0%; Pred. No. 0.59;  
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTTAG 23

Db 1 TGAGAGGCGAGTAGAAGCTTAG 23

```
RESULT 2
US-10-465-205-14
; Sequence 14, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptstI sequence from isolate no. 10
US-10-465-205-14

Query Match      86.1%; Score 19.8; DB 7; Length 1378;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 9 TGAATAATGCAGTAGAAGCTTAG 31

RESULT 3
US-10-465-205-7
; Sequence 7, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptstI sequence from isolate no. 6
US-10-465-205-7

Query Match      86.1%; Score 19.8; DB 7; Length 1379;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 9 TGAATAATGCAGTAGAAGCTTAG 31

RESULT 4
US-10-465-205-13
; Sequence 13, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
```

```
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptstI sequence from isolate no. 4
US-10-465-205-13

Query Match      86.1%; Score 19.8; DB 7; Length 1379;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 9 TGAATAATGCAGTAGAAGCTTAG 31

RESULT 5
US-10-465-205-16
; Sequence 16, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptstI sequence from isolate no. 1
US-10-465-205-16

Query Match      86.1%; Score 19.8; DB 7; Length 1379;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 6
US-10-465-205-8
; Sequence 8, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1384
```

```
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 5
US-10-465-205-8

Query Match      86.1%; Score 19.8; DB 7; Length 1384;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 10 TGAATAATGCAGTAGAAGCTTAG 32
    |||||

RESULT 7
US-10-465-205-10
; Sequence 10, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 8
US-10-465-205-10

Query Match      86.1%; Score 19.8; DB 7; Length 1384;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 9 TGAATAATGCAGTAGAAGCTTAG 31
    |||||

RESULT 8
US-10-465-205-9
; Sequence 9, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 7
US-10-465-205-9

Query Match      86.1%; Score 19.8; DB 7; Length 1385;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 10 TGAATAATGCAGTAGAAGCTTAG 32
    |||||

RESULT 9
US-10-465-205-11
; Sequence 11, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 9
US-10-465-205-11

Query Match      86.1%; Score 19.8; DB 7; Length 1390;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 24 TGAATAATGCAGTAGAAGCTTAG 46
    |||||

RESULT 10
US-10-465-205-15
; Sequence 15, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 3
US-10-465-205-15

Query Match      86.1%; Score 19.8; DB 7; Length 1393;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
Db 10 TGAATAATGCAGTAGAAGCTTAG 32
    |||||

RESULT 11
US-10-465-205-17
; Sequence 17, Application US/10465205
```

```

; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptaI sequence from isolate no. 11
US-10-465-205-17

Query Match      86.1%; Score 19.8; DB 7; Length 1447;
Best Local Similarity 91.3%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGCGCAGTAGAAAGCTTAG 32
   ||| ||| ||| ||| ||| ||| ||| |||
Db 10 TGAATATGCAGTAGAAAGCTTAG 32

RESULT 12
US-10-282-122A-38442
; Sequence 38442, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1

```

```
Best Local Similarity 90.9%; Pred. No. 1.5e+02; Indels 2; Mismatches 0; Gaps 0;
Matches 20; Conservative 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| |||||
Db 525 GAGAAGGAAGTAGAAGCCTAG 504

RESULT 15
US-10-741-600-17691/c
; Sequence 17691, Application US/10741600
; Publication No. US20050026169A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michelle et al.
; TITLE OF INVENTION: MYOCARDIAL INFARCTION, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001499
; CURRENT APPLICATION NUMBER: US/10/741,600
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 73997
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17691
; LENGTH: 81210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-741-600-17691

Query Match 81.7%; Score 18.8; DB 8; Length 81210;
Best Local Similarity 90.9%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| |||||
Db 525 GAGAAGGAAGTAGAAGCCTAG 504

RESULT 16
US-10-504-582-15/c
; Sequence 15, Application US/10504582
; Publication No. US20050176943A1
; GENERAL INFORMATION:
; APPLICANT: Yoshitake NISHIMUNE
; APPLICANT: Hiromitsu TANAKA
; APPLICANT: Masami NOZAKI
; TITLE OF INVENTION: Mouse spermatogenesis genes, mutations of male infertility-related
; TITLE OF INVENTION: and uses thereof.
; FILE REFERENCE: 2004-1256A/WMC/00653
; CURRENT APPLICATION NUMBER: US/10/504,582
; CURRENT FILING DATE: 2004-08-13
; PRIOR APPLICATION NUMBER: JP2002-36649
; PRIOR FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: JP2002-381241
; PRIOR FILING DATE: 2002-12-27
; NUMBER OF SEQ ID NOS: 183
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-504-582-15

Query Match 79.1%; Score 18.2; DB 9; Length 720;
Best Local Similarity 87.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| |||||
Db 499 TGAGAAGGCAGTAGAAGCAGAG 477

RESULT 17
US-10-473-939-5
; Sequence 5, Application US/10473939
```

```
Publication No. US20040235717A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Chawnsiang
; TITLE OF INVENTION: SUPPRESSION OF ANDROGEN RECEPTOR
; TITLE OF INVENTION: TRANSCRIPTION THROUGH NEW PATHWAYS TO AR AND AR
; FILE REFERENCE: 21108.001304
; CURRENT APPLICATION NUMBER: US/10/473,939
; CURRENT FILING DATE: 2003-10-03
; PRIOR APPLICATION NUMBER: PCT/US02/11086
; PRIOR FILING DATE: 2002-04-05
; PRIOR APPLICATION NUMBER: 60/365,060
; PRIOR FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 60/282,265
; PRIOR FILING DATE: 2001-04-06
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 160820
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:/note =
; OTHER INFORMATION: synthetic construct
US-10-473-939-5

Query Match 79.1%; Score 18.2; DB 8; Length 160820;
Best Local Similarity 87.0%; Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| |||||
Db 55080 TGAGAAGGCAGTAGGAAGCTTTG 55102

RESULT 18
US-09-925-065A-729139/c
; Sequence 729139, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 729139
; LENGTH: 522
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-729139

Query Match 77.4%; Score 17.8; DB 4; Length 522;
Best Local Similarity 90.5%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
    ||||| ||||| ||||| |||||
Db 425 TGAGAAGGCAGAGAGAGCTT 405
```

RESULT 19  
US-09-925-065A-739199/c  
; Sequence 739199, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US 60/925,065A  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 739199  
; LENGTH: 525  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-739199

Query Match 77.4%; Score 17.8; DB 4; Length 525;  
Best Local Similarity 90.5%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGCTT 21  
|||||

DB 425 TGAGAGGCGAGTGAAGCTT 405

RESULT 20  
US-09-925-065A-728631/c  
; Sequence 728631, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 728631  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-728631

Query Match 77.4%; Score 17.8; DB 4; Length 545;  
Best Local Similarity 90.5%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGCTT 21  
|||||

DB 496 TGAGAGGCGAGTGAAGCTT 476  
RESULT 21  
US-09-925-065A-69405/c  
; Sequence 69405, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 69405  
; LENGTH: 624  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-69405

Query Match 77.4%; Score 17.8; DB 4; Length 624;  
Best Local Similarity 90.5%; Pred. No. 2.5e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGCTT 21  
|||||

DB 424 TGAGAGGCGAGTGAAGCTT 404

RESULT 22  
US-09-925-065A-8982/c  
; Sequence 8982, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8982  
; LENGTH: 863  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-8982

Query Match 77.4%; Score 17.8; DB 4; Length 863;  
Best Local Similarity 90.5%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



QY 1 TCAGAAGGCAGTAGAAGCTT 21  
 |||||  
 Db 469 TGAGAAGGCAGTGGGAAGCTT 449

RESULT 23  
 US-09-925-065A-709349/c  
 ; Sequence 709349, Application US/09925065A  
 ; Publication No. US20050228172A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; Nucleotide Polymorphisms in the Human Genome  
 ; FILE REFERENCE: 108827.135  
 ; CURRENT APPLICATION NUMBER: US 09/925,065A  
 ; PRIOR FILING DATE: 2001-08-08  
 ; PRIOR APPLICATION NUMBER: US 60/243,096  
 ; PRIOR FILING DATE: 2000-10-24  
 ; PRIOR APPLICATION NUMBER: US 60/252,147  
 ; PRIOR FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/250,092  
 ; PRIOR FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/261,766  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/289,846  
 ; PRIOR FILING DATE: 2001-05-09  
 ; NUMBER OF SEQ ID NOS: 957086  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 709349  
 ; LENGTH: 1206  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-925-065A-709349

Query Match 77.4%; Score 17.8; DB 4; Length 1206;  
 Best Local Similarity 90.5%; Pred. No. 2.7e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAGAAGGCAGTAGAAGCTT 21  
 |||||  
 Db 752 TGAGAAGGCAGTAGAAGCTT 732

RESULT 24  
 US-09-925-065A-709350/c  
 ; Sequence 709350, Application US/09925065A  
 ; Publication No. US20050228172A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; Nucleotide Polymorphisms in the Human Genome  
 ; FILE REFERENCE: 108827.135  
 ; CURRENT APPLICATION NUMBER: US/09/925,065A  
 ; PRIOR FILING DATE: 2001-08-08  
 ; PRIOR APPLICATION NUMBER: US 60/243,096  
 ; PRIOR FILING DATE: 2000-10-24  
 ; PRIOR APPLICATION NUMBER: US 60/252,147  
 ; PRIOR FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/250,092  
 ; PRIOR FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/261,766  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/289,846  
 ; PRIOR FILING DATE: 2001-05-09  
 ; NUMBER OF SEQ ID NOS: 957086  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 709350  
 ; LENGTH: 1206  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-925-065A-709350

Query Match 77.4%; Score 17.8; DB 4; Length 1206;  
 Best Local Similarity 90.5%; Pred. No. 2.7e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAGAAGGCAGTAGAAGCTT 21  
 |||||  
 Db 752 TGAGAAGGCAGTAGAAGCTT 732

RESULT 25  
 US-09-925-065A-62246  
 ; Sequence 62246, Application US/09925065A  
 ; Publication No. US20050228172A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; Nucleotide Polymorphisms in the Human Genome  
 ; FILE REFERENCE: 108827.135  
 ; CURRENT APPLICATION NUMBER: US 09/925,065A  
 ; PRIOR FILING DATE: 2001-08-08  
 ; PRIOR APPLICATION NUMBER: US 60/243,096  
 ; PRIOR FILING DATE: 2000-10-24  
 ; PRIOR APPLICATION NUMBER: US 60/252,147  
 ; PRIOR FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/250,092  
 ; PRIOR FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/261,766  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/289,846  
 ; PRIOR FILING DATE: 2001-05-09  
 ; NUMBER OF SEQ ID NOS: 957086  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 62246  
 ; LENGTH: 1253  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-925-065A-62246

Query Match 77.4%; Score 17.8; DB 4; Length 1253;  
 Best Local Similarity 90.5%; Pred. No. 2.7e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAGAAGGCAGTAGAAGCTT 21  
 |||||  
 Db 109 TGAGAAGGCAGTGGGAAGCTT 129

RESULT 26  
 US-10-425-114-24004  
 ; Sequence 24004, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdong  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Screen, Steven E.  
 ; APPLICANT: Tabaska, Jack E.  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 24004  
 ; LENGTH: 1714  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB3606-055-G8\_FLI  
 US-10-425-114-24004

Query Match 77.4%; Score 17.8; DB 7; Length 1714;

Best Local Similarity 90.5%; Pred. No. 2.8e+02; Mismatches 2; Indels 0; Gaps 0;

Matches 19; Conservative 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||

Db 968 TGAGAAGGCAGTAGAAGCTT 988  
|||||

## RESULT 27

US-10-027-632-97987/c

; Sequence 97987, Application US/10027632

; Publication No. US20020198371A1

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 97987

; LENGTH: 1940

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-97987

Query Match 77.4%; Score 17.8; DB 5; Length 1940;

Best Local Similarity 90.5%; Pred. No. 2.9e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||

Db 469 TGAGAAGGCAGTAGAAGCTT 449  
|||||

## RESULT 28

US-10-027-632-97987/c

; Sequence 97987, Application US/10027632

; Publication No. US20030204075A9

; GENERAL INFORMATION:

; APPLICANT: Wang, David G.

; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

; FILE REFERENCE: 108827.129

; CURRENT APPLICATION NUMBER: US/10/027,632

; CURRENT FILING DATE: 2002-04-30

; PRIOR APPLICATION NUMBER: US 60/218,006

; PRIOR FILING DATE: 2000-07-12

; PRIOR APPLICATION NUMBER: US 60/198,676

; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: US 60/185,218

; PRIOR FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23

; PRIOR APPLICATION NUMBER: US 60/156,358

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09

; NUMBER OF SEQ ID NOS: 325720

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 97987

; LENGTH: 1940

; TYPE: DNA

; ORGANISM: Human

US-10-027-632-97987

Query Match 77.4%; Score 17.8; DB 6; Length 1940;

Best Local Similarity 90.5%; Pred. No. 2.9e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||

Db 469 TGAGAAGGCAGTAGAAGCTT 449  
|||||

## RESULT 29

US-10-425-115-112272

; Sequence 112272, Application US/10425115

; Publication No. US20040214272A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53222)B

; CURRENT APPLICATION NUMBER: US/10/425,115

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 369326

; SEQ ID NO 112272

; LENGTH: 1991

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577\_33880C.1

US-10-425-115-112272

Query Match 77.4%; Score 17.8; DB 8; Length 1991;

Best Local Similarity 90.5%; Pred. No. 2.9e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||

Db 969 TGAGAAGGCAGTAGAAGCTT 989  
|||||

## RESULT 30

US-10-437-963-95439

; Sequence 95439, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kovalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 95439

; LENGTH: 2478

; TYPE: DNA

```

; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MKT4530_93633C.1
US-10-437-963-95439

Query Match      77.4%; Score 17.8; DB 7; Length 2478;
Best Local Similarity 90.5%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
Db 1638 TGAGAAGGCAGTAGAAGCTT 1658

RESULT 31
US-10-108-260A-1480/c
; Sequence 1480, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1480
; LENGTH: 2533
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1480

Query Match      77.4%; Score 17.8; DB 6; Length 2533;
Best Local Similarity 90.5%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
Db 702 TGAGAAGGCAGTAGAAGCTT 682

RESULT 32
US-10-027-632-114128/c
; Sequence 114128, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 114128
; LENGTH: 3279
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114128

Query Match      77.4%; Score 17.8; DB 5; Length 3279;
Best Local Similarity 90.5%; Pred. No. 3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21
Db 424 TGAGAAGGCAGTAGAAGCTT 404

RESULT 34
US-10-027-632-114128/c
; Sequence 114128, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
```

; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 114128  
; LENGTH: 3279  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-114128

Query Match 77.4%; Score 17.8; DB 6; Length 3279;  
Best Local Similarity 90.5%; Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 19; Conservative 0

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
DB 424 TGAGAAGGCAGTAGAGAGCTT 404

RESULT 35  
US-10-027-632-114129/c  
; Sequence 114129, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108927.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 114129  
; LENGTH: 3279  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-114129

Query Match 77.4%; Score 17.8; DB 6; Length 3279;  
Best Local Similarity 90.5%; Pred. No. 3e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 19; Conservative 0

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
DB 424 TGAGAAGGCAGTAGAGAGCTT 404

RESULT 36  
US-11-097-143-16906  
; Sequence 16906, Application US/11097143  
; Publication No. US2005020858A1  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; TITLE OF INVENTION: DROSOPHILA GENES.

; FILE REFERENCE: CL000728  
; CURRENT APPLICATION NUMBER: US/11/097,143  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16906  
; LENGTH: 6805  
; TYPE: DNA  
; ORGANISM: DROSOPHILA  
US-11-097-143-16906

Query Match 77.4%; Score 17.8; DB 10; Length 6805;  
Best Local Similarity 90.5%; Pred. No. 3.3e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
DB 2650 TGAGAAGGCAGTAGAAGCTT 2670

RESULT 37  
US-10-741-601-5754/c  
; Sequence 5754, Application US/10741601  
; Publication No. US20040166519A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001500  
; CURRENT APPLICATION NUMBER: US/10/741,601  
; CURRENT FILING DATE: 2003-12-22  
; NUMBER OF SEQ ID NOS: 26415  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5754  
; LENGTH: 17391  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-741-601-5754

Query Match 77.4%; Score 17.8; DB 7; Length 17391;  
Best Local Similarity 90.5%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTT 21  
|||||  
DB 9559 TGAGAAGGCAGTAGAGAGCTT 9539

RESULT 38  
US-10-741-601-5677  
; Sequence 5677, Application US/10741601  
; Publication No. US20040166519A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: STENOSIS, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001500

```
; CURRENT APPLICATION NUMBER: US/10/741,601
; CURRENT FILING DATE: 2003-12-22
; NUMBER OF SEQ ID NOS: 26415
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5677
; LENGTH: 48265
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(48265)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-741-601-5677

Query Match 77.4%; Score 17.8; DB 7; Length 48265;
Best Local Similarity 90.5%; Pred. No. 4.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGCTT 21
Db 2089 TGAGAGGCGAGTGAAGCTT 2109

RESULT 39
US-10-322-281-123/c
; Sequence 123, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123
; LENGTH: 104399
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(104399)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-123

Query Match 77.4%; Score 17.8; DB 7; Length 104399;
Best Local Similarity 90.5%; Pred. No. 4.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAAGGCGAGTGAAGCTTA 22
Db 35008 GAGAAGGCGAGTGAAGACTA 34988

RESULT 40
US-10-465-205-1
; Sequence 1, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: DNA
```

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-10-465-205-1

Query Match 75.7%; Score 17.4; DB 7; Length 21;
Best Local Similarity 94.7%; Pred. No. 2.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 AAGCAGTAGAAGCTTAG 23
Db 2 AATGCAGTAGAAGCTTAG 20

RESULT 41
US-10-469-285-225/c
; Sequence 225, Application US/10469285
; Publication No. US20040126776A1
; GENERAL INFORMATION:
; APPLICANT: LION Bioscience AG
; TITLE OF INVENTION: Gene library
; FILE REFERENCE: L 1541
; CURRENT APPLICATION NUMBER: US/10/469,285
; CURRENT FILING DATE: 2003-08-28
; NUMBER OF SEQ ID NOS: 840
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 225
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Mus Musculus
; OTHER INFORMATION:
US-10-469-285-225

Query Match 75.7%; Score 17.4; DB 7; Length 309;
Best Local Similarity 94.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGC 19
Db 252 TGAGAGGCGAGCAGAAAGC 234

RESULT 42
US-09-917-800A-1065/c
; Sequence 1065, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
```

```
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1065
; LENGTH: 322
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 A1179539
US-09-917-800A-1065
```

```
Query Match          75.7%; Score 17.4; DB 3; Length 322;
Best Local Similarity 94.7%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TGAGAAGGCAGTAGAAGC 19
   |||||
Db 252 TGAGAAGGCAGCAGAAAGC 234
```

```
RESULT 43
US-09-925-065A-560212/c
; Sequence 560212, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 560212
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-560212
```

```
Query Match          75.7%; Score 17.4; DB 4; Length 569;
Best Local Similarity 94.7%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 3 AGAAGGCAGTAGAAGCTT 21
   |||||
Db 345 AAAAGGCAGTAGAAGCTT 327
```

```
RESULT 44
US-10-764-420-2470
; Sequence 2470, Application US/10764420
; Publication No. US20050084872A1
; GENERAL INFORMATION:
; APPLICANT: Lum, Pek Yee
; APPLICANT: Tan, Yejun
; APPLICANT: Dai, Hongyue
; TITLE OF INVENTION: Methods For Determining Whether An Agent
; FILE REFERENCE: ROSA122057
; CURRENT APPLICATION NUMBER: US/10/764,420
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: US 60/442,797
; PRIOR FILING DATE: 2003-01-24
```

```
; PRIOR APPLICATION NUMBER: US 60/474,413
; PRIOR FILING DATE: 2003-05-30
; NUMBER OF SEQ ID NOS: 3683
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2470
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-764-420-2470
```

```
Query Match          75.7%; Score 17.4; DB 9; Length 2151;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TGAGAAGGCAGTAGAAGC 19
   |||||
Db 1891 TGAGAAGGCAGCAGAAAGC 1909
```

```
RESULT 45
US-11-097-143-40201
; Sequence 40201, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; FILE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40201
; LENGTH: 24789
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-40201
```

```
Query Match          75.7%; Score 17.4; DB 10; Length 24789;
Best Local Similarity 94.7%; Pred. No. 5.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TGAGAAGGCAGTAGAAGC 19
   |||||
Db 13733 TGAGAAGGCAGCAGAAAGC 13751
```

```
RESULT 46
US-10-535A-45247
; Sequence 45247, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
```

```

; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 5994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45247
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Human
; US-10-242-535A-45247

```

Query Match 74.8%; Score 17.2; DB 7; Length 300;  
Best Local Similarity 86.4%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0

**Qy**

2 GAGAAAGGCAGTAGAAAAGCTTAG 23  
||| ||| ||| ||| ||| ||| ||| |||  
**pB** 137· GAGGAGGGAGTAGAAAAGCTGAG 158

```

RESULT 47
US-10-085-783A-45247
; Sequence 45247, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C.C.
; TITLE OF INVENTION: Compositions and Methods
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent version 3.2
; SEQ ID NO 45247
; LENGTH: 300

```

; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-085-783A-45247

Query Match 1 74.8%; Score 17.2; DB 7; Length 300;  
Best Local Similarity 86.4%; Pred. No. 4.4e+02;  
Matches 19: Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY            2 GAGAAGGCAGTAGAAAAGCTTAG 23  
               ||| ||| ||| ||| ||| ||| |||  
Db          137 GAGGAGGGGAGTAGAAAAGCTGAG 158

RESULT 48  
US-10-062-254-187/c  
; Sequence 187, Application US/10062254  
; Publication No. US2002013882A1  
; GENERAL INFORMATION:  
; APPLICANT: Cahoon, Edgar B  
; APPLICANT: Cahoon, Rebecca E  
; APPLICANT: Falco, Saverio Carl  
; APPLICANT: Pang, Yiwen  
; APPLICANT: Hantke, Sabine S.  
; APPLICANT: Lee, Jian-Ming

; APPLICANT: Li, Zhongsen  
 ; APPLICANT: Miao, Guo-Hua  
 ; APPLICANT: Morgante, Michele  
 ; APPLICANT: Niu, Xiping  
 ; APPLICANT: Odell, Joan  
 ; APPLICANT: Rafaleki, Antoni  
 ; APPLICANT: Sakai, Hajime  
 ; APPLICANT: Zheng, Feizhong  
 ; APPLICANT: Zhu, Qun  
 ; TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved In Plant Metabolism  
 ; FILE REFERENCE:  
 ; CURRENT APPLICATION NUMBER: US/10/062,254  
 ; CURRENT FILING DATE: 2002-02-01  
 ; PRIOR APPLICATION NUMBER: 09/630,346  
 ; PRIOR FILING DATE: 2000-07-28  
 ; PRIOR APPLICATION NUMBER: 60/146511  
 ; PRIOR FILING DATE: 1999-07-30  
 ; PRIOR APPLICATION NUMBER: 60/156006  
 ; PRIOR FILING DATE: 1999-09-23  
 ; PRIOR APPLICATION NUMBER: 60/156899  
 ; PRIOR FILING DATE: 1999-09-30  
 ; PRIOR APPLICATION NUMBER: 60/157287  
 ; PRIOR FILING DATE: 1999-10-01  
 ; PRIOR APPLICATION NUMBER: 60/169767  
 ; PRIOR FILING DATE: 1999-12-09  
 ; PRIOR APPLICATION NUMBER: 60/171054  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: 60/172958  
 ; PRIOR FILING DATE: 1999-12-21  
 ; PRIOR APPLICATION NUMBER: 60/171515  
 ; PRIOR FILING DATE: 1999-12-22  
 ; PRIOR APPLICATION NUMBER: 60/173535  
 ; PRIOR FILING DATE: 1999-12-29  
 ; NUMBER OF SEQ ID NOS: 375  
 ; SOFTWARE: Microsoft Office 97  
 ; SEQ ID NO 187  
 ; LENGTH: 494  
 ; TYPE: DNA  
 ; ORGANISM: Glycine max  
 ; FEATURE:  
 ; NAME/KEY: unsure  
 ; LOCATION: (392)  
 ; NAME/KEY: unsure  
 ; LOCATION: (409)  
 ; NAME/KEY: unsure  
 ; LOCATION: (449)  
 ; NAME/KEY: unsure  
 ; LOCATION: (464)  
 ; US-10-062-254-187

Query Match 74.8%; Score 17.2; DB 5; Length 494;  
Best Local Similarity 86.4%; Pred. No. 4.7e+02;  
Matches 19: Conservative 0; Mismatches 3; Indels

Qy 1 TGAGAAGGCAGTAGAAAGCTTA 22  
|||||  
Db 32 TGAGAAGGCAGTAGAAAGCTTA 11

```

RESULT 49
US-09-917-800A-436
; Sequence 436, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A

```

```

; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 436
; LENGTH: 519
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA965190
; NAME/KEY: misc.feature
; LOCATION: (1)..(519)
; OTHER INFORMATION: n = a or c or g or t
US-09-917-800A-436

Query Match          74.8%; Score 17.2; DB 3; Length 519;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2  GAGAGGCGAGTGAAGGCTTAG 23
Db      179 GGGAAAGCAGTAAAGGCTTAG 200

RESULT 50
US-10-027-632-237122
; Sequence 237122, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237122
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237122

Query Match          74.8%; Score 17.2; DB 6; Length 546;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  TGAGAAGGCGAGTGAAGGCTTA 22
Db      298 TGAGAAGGCGAGTAGAATATAA 319

RESULT 52
US-10-027-632-49562
; Sequence 49562, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237122
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237122
```

```

Query Match          74.8%; Score 17.2; DB 5; Length 546;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  TGAGAAGGCGAGTGAAGGCTTA 22
Db      298 TGAGAAGGCGAGTAGAATATAA 319

RESULT 51
US-10-027-632-237122
; Sequence 237122, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 237122
; LENGTH: 546
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-237122

Query Match          74.8%; Score 17.2; DB 6; Length 546;
Best Local Similarity 86.4%; Pred. No. 4.7e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  TGAGAAGGCGAGTGAAGGCTTA 22
Db      298 TGAGAAGGCGAGTAGAATATAA 319

RESULT 52
US-10-027-632-49562
; Sequence 49562, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
```



```
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49562
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49562

Query Match 74.8%; Score 17.2; DB 5; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
DB 92 GAGAAGGAGGAGAAAGCTCAG 113

RESULT 53
US-10-027-632-49563
; Sequence 49563, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49563
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49563

Query Match 74.8%; Score 17.2; DB 5; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
DB 92 GAGAAGGAGGAGAAAGCTCAG 113

RESULT 54
US-10-027-632-49564
; Sequence 49564, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49564
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49564

Query Match 74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
DB 92 GAGAAGGAGGAGAAAGCTCAG 113

RESULT 55
US-10-027-632-49562
; Sequence 49562, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49562
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49562

Query Match 74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
DB 92 GAGAAGGAGGAGAAAGCTCAG 113
```

```
RESULT 56
US-10-027-632-49563
; Sequence 49563, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49563
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49563

Query Match      74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAAGCTTAG 23
DB      92 GAGAAGGAAGGAGAAAGCTCAG 113

RESULT 57
US-10-027-632-49564
; Sequence 49564, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49564
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49564

Query Match      74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAAGCTTAG 23
DB      92 GAGAAGGAAGGAGAAAGCTCAG 113

RESULT 58
US-09-925-065A-34
; Sequence 34, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-34

Query Match      74.8%; Score 17.2; DB 4; Length 583;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAAGCTTAG 23
DB      249 GAAAGGCATTAGAAAGCATAG 270

RESULT 59
US-10-972-079-87189
; Sequence 87189, Application US/10972079
; Publication No. US2005015317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; CATTLE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 87189
; LENGTH: 600
; TYPE: DNA
```

```
; ORGANISM: Human
US-10-027-632-49564

Query Match      74.8%; Score 17.2; DB 6; Length 568;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAAGCTTAG 23
DB      92 GAGAAGGAAGGAGAAAGCTCAG 113

RESULT 58
US-09-925-065A-34
; Sequence 34, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-34

Query Match      74.8%; Score 17.2; DB 4; Length 583;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAAGCTTAG 23
DB      249 GAAAGGCATTAGAAAGCATAG 270

RESULT 59
US-10-972-079-87189
; Sequence 87189, Application US/10972079
; Publication No. US2005015317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF
; CATTLE
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 87189
; LENGTH: 600
; TYPE: DNA
```

```
; ORGANISM: Chicken 19866894388660_1
US-10-972-079-87189

Query Match          74.8%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAGCTTAG 23
DB      469 GAAAGGCGAGCAGAAAGCTGAG 490

RESULT 60
US-10-972-079-87190
; Sequence 87190, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEER
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 87190
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894388660_2
US-10-972-079-87190

Query Match          74.8%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 4.8e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAGGCGAGTAGAAGCTTAG 23
DB      266 GAAAGGCGAGCAGAAAGCTGAG 287
```

Search completed: January 28, 2006, 01:45:28  
Job time : 405.011 secs

**This Page Blank (uspto)**

ECOR1 adaptor) (Start Insert). . . . .AAAAAAAAAAAAAAAAAAAA (End Insert) (Start Tag)TGGCA(End Tag) (Start NotI site/Vector)GGCGCCGCCACCGG. . . The total number of white colony forming units (cfu) in the primary library before amplification was 2.1x10<sup>6</sup> cfu (colony forming units). The background of empty clones was less than 1%. Inserts ranged from 0.5kb to 4 kb, as determined by PCR. Purified plasmid DNA from the primary library was converted to single-stranded circles and used as a template for PCR amplification using the T7 and T3 priming sites flanking the cloned cDNA inserts. The purified PCR products, representing the entire cloned cDNA population, were used as a driver for normalization. Hybridization between the single-stranded library and the PCR products was carried out for 44 hours at 30C. Unhybridized single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B cells to generate the normalized library. The total number of clones with insert was 5.6x10<sup>6</sup> cfu. Background of empty clones was less than 1%."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 7; Length 367;  
Best Local Similarity 90.5%; Pred. No. 3.2e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTAG 23  
||||| ||||| ||||| |||||  
Db 130 AGAAGGCTGTAGAATCTTAG 150

## RESULT 41

BE142442  
LOCUS RC2-HT0148-071099-013-h01 HT0148 Homo sapiens cDNA, mRNA linear EST 21-JUN-2000  
DEFINITION BE142442  
ACCESSION BE142442.1 GI:8605163  
VERSION  
KEYWORDS EST.  
SOURCE Homo sapiens (human)

## ORGANISM

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 428)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Mateukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=RC2-HT0148-071099-013-h01&t3=1999-10-07&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 16  
High quality sequence stop: 142.  
Location/Qualifiers  
1. 428  
/organism="Homo sapiens"  
/mol\_type="mRNA"

## FEATURES

source  
1. 428  
/organism="Homo sapiens"  
/mol\_type="mRNA"

/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="HT0148"

/note="Organ: head neck; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 2; Length 428;  
Best Local Similarity 90.5%; Pred. No. 3.3e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTAG 23  
||||| ||||| ||||| |||||  
Db 375 AGAAGGCAGTAGACAGCTTAG 395

## RESULT 42

BM304832  
LOCUS SNEST4a32h03.y1 cSn 1 S neuropa invitro merozoite cDNA Sarcocystis  
DEFINITION BM304832  
ACCESSION BM304832.1 GI:18036536  
VERSION  
KEYWORDS EST.  
SOURCE Sarcocystis neuropa  
ORGANISM Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Sarcocystis.

## REFERENCE

1 (bases 1 to 442)  
Howe, D.K., Stamper, S., Tang, K., Sibley, L.D., Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, F., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., McCann, R., Blistain, A., Bennet, J., Schmitt, A., Ronko, I., Tsagareishvili, R., Fedele, M., Belaygorod, L., Franklin, C., Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R. and Wilson, R.  
Sarcocystis neuropa EST project  
Unpublished (2000)  
Contact: Daniel K. Howe  
Sarcocystis neuropa EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 347.  
Location/Qualifiers  
1. 442  
/organism="Sarcocystis neuropa"  
/mol\_type="mRNA"  
/strain="Sn3"  
/db\_xref="taxon:42890"  
/dev\_stage="merozoite"  
/lab\_host="DH10B"

## TITLE

JOURNAL  
COMMENT  
Sarcocystis neuropa EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 347.  
Location/Qualifiers  
1. 442  
/organism="Sarcocystis neuropa"  
/mol\_type="mRNA"  
/strain="Sn3"  
/db\_xref="taxon:42890"  
/dev\_stage="merozoite"  
/lab\_host="DH10B"

## FEATURES

source  
1. 442  
/organism="Sarcocystis neuropa"  
/mol\_type="mRNA"  
/strain="Sn3"  
/db\_xref="taxon:42890"  
/dev\_stage="merozoite"  
/lab\_host="DH10B"  
/clone\_lib="cSn 1 S neuropa invitro merozoite cDNA"  
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; The library was constructed by Dan Howe, University of Kentucky. cDNAs were synthesized from poly(A)+ RNA by oligo d(T) priming and directionally cloned into the Uni-ZAP XR lambda vector. The library was mass excised as phagemids and rescued in SOLR cells. The plasmid library was recovered from the SOLR cells and transformed in mass into DH10B cells for sequencing. WARNING: the library contains a small percentage of cDNAs derived from the bovine host cells."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 3; Length 442;  
 Best Local Similarity 90.5%; Pred. No. 3.3e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGGCTTA 22  
 |||||  
 DB 201 GAGAGGCGAGTAGAAGGCTTA 221

RESULT 43  
 BF323664  
 LOCUS  
 DEFINITION SNEST4a18a11.y1 csn 1 S neurona invitro merozoite cDNA Sarcocystis  
 neurona cDNA 5', mRNA sequence.  
 ACCESSION BF323664  
 VERSION BF323664  
 KEYWORDS EST.  
 SOURCE Sarcocystis neurona  
 ORGANISM Sarcocystis neurona

REFERENCE 1 (bases 1 to 454)  
 Sarcocystidae; Apicomplexa; Coccidia; Eimeriida;  
 Sarcocystis.

AUTHORS Howe, D.K., Stamper, S., Tang, K., Sibley, L.D., Clifton, S., Marra, M.,  
 Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B., Bowers, Y.,  
 Gibbons, M., Ritter, E., McCann, R., Blistain, A., Bennett, J.,  
 Schmitt, A., Ronko, I., Tsagareishvili, R., Fedele, M., Belaygorod, L.,  
 Franklin, C., Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J.,  
 Waterston, R. and Wilson, R.  
 Sarcocystis neurona EST project  
 Unpublished (2000)  
 Contact: Daniel K. Howe  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further  
 information relating to organism, libraries, or clone availability.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 415.

TITLE Sarcocystis neurona EST project  
 JOURNAL  
 COMMENT

FEATURES  
 source  
 1..454  
 /organism="Sarcocystis neurona"  
 /mol\_type="mRNA"  
 /strain="Sn3"  
 /db\_xref="taxon:42890"  
 /dev\_stage="merozoite"  
 /lab\_host="DH10B"  
 /clone\_lib="csn 1 S neurona invitro merozoite cDNA"  
 /notes="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:  
 xhoI; The library was constructed by Dan Howe, University  
 of Kentucky. cDNAs were synthesized from poly(A) + RNA  
 by oligo d(T) priming and directionally cloned into the  
 Uni-ZAP XR lambda vector. The library was mass excised  
 as phagemids and rescued in SOLR cells. The plasmid  
 library was recovered from the SOLR cells and transformed  
 in mass into DH10B cells for sequencing. WARNING: the  
 library contains a small percentage of cDNAs derived from  
 the bovine host cells."

ORIGIN  
 Query Match 77.4%; Score 17.8; DB 2; Length 454;  
 Best Local Similarity 90.5%; Pred. No. 3.3e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGGCTTA 22  
 |||||  
 DB 196 GAGAGGCGAGTAGAAGGCTTA 216

RESULT 44

BM305256  
 LOCUS  
 DEFINITION SNEST4a38c04.y1 csn 1 S neurona invitro merozoite cDNA Sarcocystis  
 neurona cDNA 5', mRNA sequence.  
 ACCESSION BM305256  
 VERSION BM305256.1  
 KEYWORDS GI:18036960  
 SOURCE EST.  
 ORGANISM Sarcocystis neurona

REFERENCE 1 (bases 1 to 501)  
 Sarcocystidae; Apicomplexa; Coccidia; Eimeriida;  
 Sarcocystis.

AUTHORS Howe, D.K., Stamper, S., Tang, K., Sibley, L.D., Clifton, S., Marra, M.,  
 Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B., Bowers, Y.,  
 Gibbons, M., Ritter, E., McCann, R., Blistain, A., Bennett, J.,  
 Schmitt, A., Ronko, I., Tsagareishvili, R., Fedele, M., Belaygorod, L.,  
 Franklin, C., Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J.,  
 Waterston, R. and Wilson, R.  
 Sarcocystis neurona EST project  
 Unpublished (2000)  
 Contact: Daniel K. Howe  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu  
 Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further  
 information relating to organism, libraries, or clone availability.  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 425.

TITLE Sarcocystis neurona EST project  
 JOURNAL  
 COMMENT

FEATURES  
 source  
 1..501  
 /organism="Sarcocystis neurona"  
 /mol\_type="mRNA"  
 /strain="Sn3"  
 /db\_xref="taxon:42890"  
 /dev\_stage="merozoite"  
 /lab\_host="DH10B"  
 /clone\_lib="csn 1 S neurona invitro merozoite cDNA"  
 /note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:  
 xhoI; The library was constructed by Dan Howe, University  
 of Kentucky. cDNAs were synthesized from poly(A) + RNA  
 by oligo d(T) priming and directionally cloned into the  
 Uni-ZAP XR lambda vector. The library was mass excised  
 as phagemids and rescued in SOLR cells. The plasmid  
 library was recovered from the SOLR cells and transformed  
 in mass into DH10B cells for sequencing. WARNING: the  
 library contains a small percentage of cDNAs derived from  
 the bovine host cells."

ORIGIN

Query Match 77.4%; Score 17.8; DB 3; Length 501;  
 Best Local Similarity 90.5%; Pred. No. 3.3e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGGCTTA 22  
 |||||  
 DB 122 GAGAGGCGAGTAGAAGGCTTA 142

RESULT 45  
 AZ851239/c  
 LOCUS  
 DEFINITION 2M0153G22F Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC2M0153G22 F, genomic survey sequence.

ACCESSION AZ851239  
 VERSION AZ851239.1  
 KEYWORDS GI:13037038  
 SOURCE GSS.  
 ORGANISM Mus musculus  
 (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

```

REFERENCE
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Becorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.
TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCT, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0153 row: G column: 22
Seq primer: CGTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 502.
Location/Qualifiers
FEATURES
source
1. .502
/organism="Mus musculus"
/mol_type="genomic DNA"
/bstrain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0153G22"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC11 library"
/notes="Vector: FWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydronamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pW42 [gi|4732114|gb|AF129072.1], a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 77.4%; Score 17.8; DB 9; Length 502;
Best Local Similarity 90.5%; Pred. No. 3.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAAGGCAGTAGAAGCTTA 22
|||||
Db 315 GAGAAGGCAGTAGAAGACTA 295

RESULT 46
AQ437381
LOCUS
DEFINITION
HS 5122 A1.D03.T7A.RP11-11 Human Male BAC Library Homo sapiens
Genomic_clone Plate=698 Col=5 Row=G, genomic survey sequence.
ACCESSION
AQ437381
VERSION
AQ437381.1 GI:4548720
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Sciurognathi; Muroidea; Muridae; Murinae; Mus.
1 (bases 1 to 502)
Dunn,D., Aoyagi,A., Barber,M., Becorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niederhausern,A. and Wright,D.,Weiss,R.
TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL
Unpublished (2000)
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLCT, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0153 row: G column: 22
Seq primer: CGTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 502.
Location/Qualifiers
FEATURES
source
1. .502
/organism="Mus musculus"
/mol_type="genomic DNA"
/bstrain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0153G22"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUGC11 library"
/notes="Vector: FWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydronamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pW42 [gi|4732114|gb|AF129072.1], a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN
Query Match 77.4%; Score 17.8; DB 9; Length 502;
Best Local Similarity 90.5%; Pred. No. 3.3e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGAAGGCAGTAGAAGCTTA 22
|||||
Db 315 GAGAAGGCAGTAGAAGACTA 295

RESULT 46
AQ437381
LOCUS
DEFINITION
HS 5122 A1.D03.T7A.RP11-11 Human Male BAC Library Homo sapiens
Genomic_clone Plate=698 Col=5 Row=G, genomic survey sequence.
ACCESSION
AQ437381
VERSION
AQ437381.1 GI:4548720
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominiidae; Homo.
1 (bases 1 to 522)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
COMMENT
10449764
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 698 row: G column: 5
Seq primer: T7
Class: BAC ends
High quality sequence stop: 522.
Location/Qualifiers
FEATURES
source
1. .522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=698 Col=5 Row=G"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/notes="Vector: pBACe3.6; Site:1: EcoRI; Site:2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

ORIGIN
Query Match 77.4%; Score 17.8; DB 9; Length 522;
Best Local Similarity 90.5%; Pred. No. 3.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGCGCAGTAGAAGCTT 21
|||||
Db 393 TGAGAAGCGCAGTAGAGCTT 413

RESULT 47
BM303377
LOCUS
DEFINITION
SNEST4a58c03.y1 cSN 1 S neurona invitro merozoite cDNA Sarcocystis
neurona cDNA 5', mRNA sequence.
ACCESSION
BM303377
VERSION
BM303377.1 GI:18035081
KEYWORDS
EST.
SOURCE
Sarcocystis neurona
ORGANISM
Sarcocystis neurona
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Sarcocystis.
1 (bases 1 to 534)
Howe,D.K., Stamper,S., Tang,K., Sibley,L.D., Clifton,S., Marra,M.,
Hillier,L., Pape,D., Martin,J., Wyllie,T., Theising,B., Bowers,T.,
Gibbons,M., Ritter,E., McCann,R., Blistain,A., Bannat,J.,
Schmitt,A., Ronko,I., Tsagareishvili,R., Fedele,M., Belaygorod,L.,
Franklin,C., Carr,L.M., Grow,A., Maguire,L., Wadkins,J., Richey,J.,
Waterston,R. and Wilson,R.
Sarcocystis neurona EST project
Unpublished (2000)
Contact: Daniel K. Howe
TITLE
Sarcocystis neurona EST project
JOURNAL
Unpublished (2000)
COMMENT
Contact: Daniel K. Howe

```

Sarcocystis neuropa EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
Contact Daniel K. Howe (dkhowe2@pop.uky.edu) for further  
information relating to organism, libraries, or clone availability.  
Seq primer: -40RP from Gbco  
High quality sequence stop: 420.

## FEATURES

source

```
1. .534
/organism="Sarcocystis neuropa"
/mol_type="mRNA"
/strain="Sn3"
/db_xref="taxon:42890"
/dev_stage="merozoite"
/lab_host="DH10B"
/clone_lib="cSn 1 S neuropa invitro merozoite cDNA"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; The library was constructed by Dan Howe, University
of Kentucky. cDNAs were synthesized from poly(A)+ RNA
by oligo d(T) priming and directionally cloned into the
Uni-ZAP XR lambda vector. The library was mass excised
as phagemids and rescued in SOLR cells. The plasmid
library was recovered from the SOLR cells and transformed
in mass into DH10B cells for sequencing. WARNING: the
library contains a small percentage of cDNAs derived from
the bovine host cells."
```

## ORIGIN

```
Query Match 77.4%; Score 17.8; DB 3; Length 534;
Best Local Similarity 90.5%; Pred. No. 3.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAAAGCTTA 22
|||||
DB 2 GAGAGGCGAGTAAAGCTTA 22
```

## RESULT 48

```
AQ595160/c
LOCUS
DEFINITION
HS_5424_B1_G09_T7A RPCI-11 Human Male BAC Library Homo sapiens
Genomic clone Plate=1000 Col=17 Row=N, Genomic survey sequence.
ACCESSION
AQ595160
VERSION
GSS.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
```

## REFERENCE

```
1 (bases 1 to 537)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
```

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome

Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

## JOURNAL

PUBMED

10449764

## COMMENT

Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Dieter de Jong  
(pieterdejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm)  
or from Resear h Genetics (info@resgen.com). BAC end Web Server:

http://www.htac.washington.edu  
Plate: 1000 row: N column: 17  
Seq primer: 17  
Class: BAC ends  
High quality sequence stop: 537.

## FEATURES

source

```
1. 537
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=1000 Col=17 Row=N"
/sex="male"
/clone_lib="RPCI-11 Human Male BAC Library"
/notes="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
```

## ORIGIN

```
Query Match 77.4%; Score 17.8; DB 9; Length 537;
Best Local Similarity 90.5%; Pred. No. 3.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGCGAGTAGAAGCTT 21
|||||
DB 438 TGTGAAGCGAGTAGAAGCTT 418
```

## RESULT 49

```
CD313569
LOCUS
DEFINITION
CD313569
ACCESSION
CD313569
VERSION
EST.
KEYWORDS
Strongylocentrotus purpuratus
SOURCE
Strongylocentrotus purpuratus
ORGANISM
```

## REFERENCE

```
1 (bases 1 to 567)
Poustka,A.J., Groth,D., Hennig,S., Thamm,S., Cameron,A., Beck,A.,
Reinhardt,R., Herwig,R., Panopoulou,G. and Lehrach,H.
Generation, annotation, evolutionary analysis, and database
integration of 20,000 unique sea urchin EST clusters
Genome Res. 13 (12), 2736-2746 (2003)
```

## JOURNAL

COMMENT

Contact: Poustka AJ  
Laboratory 145, dept.Lehrach  
Max-Planck-Institut fuer Molekulare Genetik  
Innestr.63-73, D-14195 Berlin, Germany  
Tel: +49 30 8413 1235  
Fax: +49 30 8413 1128  
Email: poustka@molgen.mpg.de  
The library was characterised by oligonucleotide fingerprinting  
(ONF) to reduce sequencing redundancy. According to the ONF  
procedure, clones that display the same hybridisation matrix with a  
battery of 200 8mer oligonucleotides are grouped into clusters. One  
clone per ONF cluster is selected for sequencing. The size of each  
cluster is an indicator of the frequency of a transcript in the  
analysed library. The cluster size as well as the coordinates of  
the other clones assigned to the same ONF cluster as the clone from  
which the above EST is generated is available at the sea urchin  
project web site at: http://www.molgen.mpg.de/ag\_seaurchin/. cDNA  
clones and filters are distributed via the Resource Center/Primary  
Database of the German Human Genome Project (http://www.rzpd.de)  
PCR Primers  
FORWARD: 5' CCCAGGCTTTACACTTATGCTTCGGCTCG 3' (M13RSP) 5'-seq  
BACKWARD: 5' GCTATTACCGCAGTCGGGAAGGGGATG 3' (M13FSP) 3'-seq  
Seq primer: 5'-CCGTCGCGAATTCGCGGT-3' pSpout3/86  
High quality sequence stop: 567.



```

FEATURES
  source
    Location/Qualifiers
      1..567
        /organism="Strongylocentrotus purpuratus"
        /mol_type="mRNA"
        /db_xref="taxon:7668"
        /clone="MPMGp621P1813.MPI 621.13P18"
        /issue_type="whole unfertilised eggs"
        /dev_stage="embryonic 0hr"
        /lab_host="S.colli, XLI blue"
        /clone_lib="Sea urchin unfertilised egg cDNA library
        MPMGp621"
        /note="Vector: pSport1; Site 1: NotI; Site 2: SalI; Random
        primed and directionally cloned in pSport1 vector using a
        NotI (5'-P-G-A-C-T-A-G-T-T-C-A-T-C-G-C-G-C-G-C-C (T)15-3' and a
        SalI 5'-T-C-G-A-C-C-C-A-C-G-G-T-C-G-3'adapters (Gibco BRL)"

ORIGIN
  Query Match 77.4%; Score 17.8; DB 6; Length 567;
  Best Local Similarity 90.5%; Pred. No. 3.4e+03;
  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGCAGTAGAAGCTT 21
   ||||| ||||| ||||| |||||
Db 29 TGAGGAGCGCAGTCGAAAGCTT 49

RESULT 50
CR221772/c
LOCUS
DEFINITION
  CR221772 579 bp DNA linear GSS 06-JUL-2004
  Reverse strand read from insert in 5'hpRT insertion targeting and
  chromosome engineering clone MHPN262d10, genomic survey sequence.

ACCESSION
  CR221772.1 GI:50000621
VERSION
  GSS; genome survey sequence; MICER.
KEYWORDS
  Mus musculus (house mouse)
SOURCE
  Mus musculus
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
    Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE
  1 (bases 1 to 579)
  Adams,D.J., Biggs,P.J., Cox,A.V., Davies,R.M., van der Weyden,L.,
  Jonkers,J., Smith,J., Plumb,R.W., Taylor,R.G., Nishijima,I., Yu,Y.,
  Rogers,J. and Bradley,A.
  Direct Submission
  Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire,
  CB10 1SA, UK. http://www.sanger.ac.uk/MICER

FEATURES
  source
    Location/Qualifiers
      1..579
        /organism="Mus musculus"
        /mol_type="genomic DNA"
        /db_xref="taxon:10090"
        /clone="MHPN262d10"
        /clone_lib="MHPN"

ORIGIN
  Query Match 77.4%; Score 17.8; DB 11; Length 579;
  Best Local Similarity 90.5%; Pred. No. 3.4e+03;
  Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCGCAGTAGAAGCTTAG 23
   ||||| ||||| ||||| |||||
Db 210 AGAAGGAAGTAGAAGAGATTAG 190

RESULT 51
CN496517
LOCUS
DEFINITION
  CN496517 585 bp mRNA linear EST 24-MAY-2004
  Mdfw2022g09.y1 Mdfw Malus x domestica cDNA clone Mdfw2022g09 5',
  mRNA sequence.
ACCESSION
  CN496517
VERSION
  CN496517.1 GI:46599127
KEYWORDS
  EST.
SOURCE
  Malus x domestica

```

## ORGANISM

Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.

REFERENCE  
AUTHORS

1 (bases 1 to 585)  
Korban,S., Vodkin,L., Liu,L., Gasic,K., Gonzales,O., Hernandez,A.,  
Aldwinckle,H., Malnoy,M., Carroll,N., Goldsbrough,P., Orville,K.,  
Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T.,  
Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Ronko,I.,  
Tsagarisvilli,R., Kennedy,S., Waterston,R. and Wilson,R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)

TITLE  
JOURNAL  
COMMENT

Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu  
Library materials provided by: Schuyler S. Korban Library  
constructed by: A. Hernandez / K. Gasic Library sequenced by:  
Washington University Genome Sequencing Center  
WashU EST name: aaf85d05.y1  
Seq primer: -40UP from Gibco.  
Location/Qualifiers

## FEATURES

source

```

1..585
  /organism="Malus x domestica"
  /mol_type="mRNA"
  /cultivar="GoldRush"
  /db_xref="taxon:3750"
  /clone="Mdfw2022g09"
  /issue_type="Flower"
  /lab_host="DH10B ampicillin resistant"
  /clone_lib="Mdfw"
  /note="Vector: pBluescript II SK (+); Site 1: NotI;
  Site 2: EcoRI; Total RNA was extracted separately from
  each floral stage (bud, balloon, open and after
  pollination), using the 'pine tree' method. Poly(A)+mRNA
  was isolated twice from total RNA from each stage using
  the Oligotex Direct mRNA kit (Qiagen). mRNA was reverse
  transcribed into double stranded cDNA using a modified
  oligo18(dT) primer with an identifying tag sequence (see
  table below). cDNA's from different stages were pooled in
  equal amounts before adaptor ligation. Tag identification
  when sequencing from 5' end: Stage 1 (bud) insert
  18(A)TCGGA; Stage 2 (balloon) insert 18(A)TCGA; Stage 3
  (open) insert 18(A)TCGT; Stage 4 (afterpollination)
  insert 18(A)TCGT. Tag identification when sequencing from
  3' end: Stage 1 (bud) TCCGA18(T) insert; Stage 2 (balloon)
  TCGCA18(T) insert; Stage 3 (open) ACGCA18(T) insert; Stage
  4 (afterpollination) ACCGA18(T) insert. Double stranded
  cDNAs were size selected (more than 450 bp), adaptor
  with EcoRI adapters at both ends and then digested with
  NotI. The cDNAs were then directionally cloned into
  EcoRI-NotI digested pBS II SK(+) phagemid
  vector(Stratagene). Identification of adaptors and tags in
  5'-end sequenced clones: (Vector) . . . TAAGCTT(End
  Vector)(Start EcoRI adaptor)GATATCGATTCTCATGCTTGGG (End
  EcoRI adaptor)(Start insert) . . . AAAAAAAAAAAAAAAAAA(End
  insert) (Start Tag)TCGA(End Tag)(Start
  NotI site/Vector)GCGCGCCGCCCGCGG. . . The total number of
  white colony forming units (cfu) in the primary library
  before amplification was 1.1x106 cfu (colony forming
  units). The background of empty clones was less than 1%.
  Inserts ranged from 0.5kb to 3 kb, as determined by PCR.
  Purified plasmid DNA from the primary library was
  converted to single-stranded circles and used as a template
  for PCR amplification using the T7 and T3 priming sites
  flanking the cloned cDNA inserts. The purified PCR
  products, representing the entire cloned cDNA population,
  were used as a driver for normalization. Hybridization
  between the single-stranded library and the PCR products
  was carried out for 44 hours at 30C. Unhybridized

```

single-stranded DNA circles were separated from hybridized DNA rendered partially double-stranded and electroporated into DH10B cells to generate the normalized library. The total number of clones with insert was 9x10<sup>6</sup> cfu. Background of empty clones was less than 1%."

```

ORIGIN
Query Match          77.4%; Score 17.8; DB 7; Length 585;
Best Local Similarity 90.5%; Pred. No. 3.4e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 AGAAGGCTAGAAAGCTTAG 23
|||||
Db 462 AGAAGGCTAGAAATCTTAG 482
|||||

RESULT 52
CB638470
LOCUS          624 bp mRNA linear EST 08-APR-2003
DEFINITION    OSJNEa06M21.f OSJNEa Oryza sativa (japonica cultivar-group) cDNA
ACCESSION    CB638470
VERSION      CB638470
KEYWORDS     EST.
SOURCE       CB638470.1 GI:29633461
ORGANISM     Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE    1 (bases 1 to 624)
AUTHORS      Jantaseuriyarat,C., Gowda,M., Haller,K., Hatfield,J., Lu,G.,
Stahlberg,E., Zhou,B., Li,H., Kim,H., Yu,Y., Dean,R.A., Wing,R.A.,
Soderlund,C. and Wang,G.-L.
TITLE        Large-scale identification of expressed sequence tags involved in
JOURNAL      Plant Physiol. 138 (1), 105-115 (2005)
PUBMED      15888683
COMMENT      Contact: Rod Wing
              Arizona Genomics Institute
              University of Arizona
              Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
              85721-0088, USA
              Tel: 520 626 3967
              Fax: 520 621 3288
              Email: rwing@genome.arizona.edu
PCR Primers  FORWARD: atc agc ggc cgc gat cc
              BACKWARD: aat taa ccc tca cta aag gg
              Plate: 06 row: M column: 21
              Seq primer: atc agc ggc cgc gat cc.
FEATURES
source
1. .624
   /organism="Oryza sativa (japonica cultivar-group)"
   /mol_type="mRNA"
   /cultivar="Nipponbare"
   /db_xref="taxon:39947"
   /clone="OSJNEa06M21"
   /tissue_type="Leaf"
   /dev_stage="3 week"
   /lab_host="DH10B"
   /clone_lib="OSJNEa"
   /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
XhoI; 6 hrs after inoculation with Rice Blast (Che
86061)"

ORIGIN
Query Match          77.4%; Score 17.8; DB 6; Length 624;
Best Local Similarity 90.5%; Pred. No. 3.5e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTT 21
|||||
Db 14 TGAGAAGGCATAGATGCTT 34
|||||

RESULT 53
CD278427
LOCUS          630 bp mRNA linear EST 01-SEP-2003
DEFINITION    T143C02371F (FHIG:C) Axenic plate culture Betula pendula cDNA 5',
mRNA sequence.
ACCESSION    CD278427
VERSION      CD278427.1 GI:34390473
KEYWORDS     EST.
SOURCE       Betula pendula (European white birch)
ORGANISM     Betula pendula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Fagales; Betulaceae; Betula.
REFERENCE    1 (bases 1 to 630)
AUTHORS      Johansson,T., Le Quere,A., Ahren,D., Soderstrom,B., Erlandsson,R.,
Lundeberg,J., Uhlen,M. and Tunlid,A.
TITLE        Transcriptional responses of Paxillus involutus and Betula pendula
JOURNAL      Mol. Plant Microbe Interact. 17 (2), 202-215 (2004)
PUBMED      14964534
COMMENT      Contact: Johansson,T.
              Fungal-Host Interaction Group (FHIG)
              Microbial Ecology, Institution of Ecology
              Ecology Building, Lund University, SE-223 62 Lund, Sweden
              Tel: +46 46 222 45 49
              Fax: +46 46 222 41 58
              Email: tomas.johansson@mbioekol.lu.se
PCR Primers  FORWARD: P104 (5'-GGGAGCGCGCCATTGTGT-3')
              BACKWARD: P105 (5'-AGTGAGCTCGAATTCGGCC-3')
              Seq primer: P104
              High quality sequence stop: 630.
FEATURES
source
1. .630
   /organism="Betula pendula"
   /mol_type="mRNA"
   /strain="Skuleskogen, Skogsforsk, Sweden"
   /db_xref="taxon:3505"
   /tissue_type="Root tissue"
   /dev_stage="25 days of growth after transfer"
   /lab_host="Escherichia coli BM25.8"
   /clone_lib="(FHIG:C) Axenic plate culture"
   /note="Vector: pTriplex2; Site 1: SfiI; Site 2: SfiI; This
EST clone is originating from one of three cDNA libraries,
constructed for transcript profiling of the mycorrhizal
interaction between the basidiomycete Paxillus involutus
and Betula pendula (birch). One library represents the
developed and functional mycorrhizal root tissue
(' (FHIG:A) Ectomycorrhiza plate culture'), a second
library represents axenically grown fungus (' (FHIG:B)
Axenic plate culture') and a third library represents
axenically grown plants (' (FHIG:C) Axenic plate culture').
Libraries were analyzed in parallel and 3555 (FHIG:A),
3964 (FHIG:B), and 2532 (FHIG:C) high-quality (PHRED 20)
ESTs of >99bp have been deposited. The cDNA libraries were
constructed from total RNA using the SMART cDNA library
Construction kit (#K1051-1, Clontech, Palo Alto, CA, USA)
according to the manufacturer's instructions. Full-length
cDNAs were trimmed by SfiI, fractionated and directionally
ligated into (lambda)Triplex2 arms. The lambda library was
converted to a plasmid library via site-specific
recombination at loxp sites in a Cre+ strain (E. coli
BM25.8). Plasmid clones were randomly collected and
analysed by DNA sequencing using a plasmid-specific
forward primer (P104)."
```

Query Match 77.4%; Score 17.8; DB 6; Length 630;

Best Local Similarity 90.5%; Pred. No. 3.5e+03;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;



Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0518 row: J column: 21  
 Seq primer: CACACAGGAACAGCTATGACC  
 Class: plasmid ends  
 High quality sequence stop: 665.  
 Location/Qualifiers  
 1. .665

FEATURES  
 source

/organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGCIM0518J21"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGCIM library"  
 /notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (GI|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 77.4%; Score 17.8; DB 9; Length 665;  
 Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTAG 23  
 Db 41 AGAAGGAAGTAGAAGATTAG 61

RESULT 57  
 CE179177/c

LOCUS 668 bp DNA linear GSS 25-SEP-2003  
 DEFINITION tigr-gss-dog-17000326760515 Dog Library Canis familiaris genomic, genomic survey sequence.

ACCESSION CE179177

VERSION CE179177.1 GI:35327228

KEYWORDS GSS.

SOURCE Canis familiaris (dog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae; Canis.

REFERENCE 1 (bases 1 to 668)

AUTHORS Kirkness,E.F., Batna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.

TITLE The dog genome: survey sequencing and comparative analysis

JOURNAL Science 301 (5641), 1898-1903 (2003)

PUBMED 14512627

COMMENT

Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-0200

Fax: 301-838-0208  
 Email: ekirknes@tigr.org  
 Class: shotgun.

FEATURES  
 source

Location/Qualifiers  
 1. .668  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 77.4%; Score 17.8; DB 9; Length 668;  
 Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGGAAGGCAGTAGAAGCTT 21

Db 266 TGAGGAAGGCAGGAGAGATT 246

RESULT 58

AQ541187/c

LOCUS 689 bp DNA linear GSS 19-MAY-1999

DEFINITION RPCI-11-344K16.TJ RPCI-11 Homo sapiens genomic clone

ACCESSION AQ541187

VERSION AQ541187.1 GI:4871643

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 689)

AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready

JOURNAL Map Building

COMMENT Other GSSs: RPCI-11-344K16.TV

Contact: Shaving Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbs@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html.

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. .689

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="GBS:7631967"

/db\_xref="taxon:9606"

/clone="RPCI-11-344K16"

/sex="Male"

/cell\_type="Lymphocytes"

/clone\_lib="RPCI-11"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;

RPCI11 Human Male BAC Library"

ORIGIN

Query Match 77.4%; Score 17.8; DB 9; Length 689;

Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGCTTA 22  
|||||  
DB 404 GAGAAAGTAGTAGAAGCTTA 384  
|||||

RESULT 59  
DRI08351/c  
LOCUS  
DEFINITION JHUI62A06L704 Canine cardiovascular system biased cDNA Canis  
familiaris cDNA, mRNA sequence.  
704 bp mRNA linear EST 13-JUN-2005

ACCESSION DRI08351  
VERSION DRI08351.1 GI:67567706  
KEYWORDS EST.  
SOURCE Canis familiaris (dog)  
ORGANISM Canis familiaris  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
Canis.  
1 (bases 1 to 704)

REFERENCE DiSilvestre, D., Yung, C., Gao, Z., Farukhi, Y., Winslow, R. L. and  
Tomaselli, G. F.  
Canine cardiovascular system biased cDNA sequences  
Contact: Gordon F. Tomaselli  
Johns Hopkins University  
720 Rutland Avenue/Ross 844, Baltimore, MD 21205, USA  
Tel: 4109552774  
Fax: 4105022096  
Email: gtomasel@jhmi.edu

FEATURES  
source  
Location/Qualifiers  
1..704  
/organism="Canis familiaris"  
/mol\_type="mRNA"  
/strain="Mixed"  
/db\_xref="taxon:9615"  
/sex="Mixed"  
/clone\_lib="Canine cardiovascular system biased cDNA"  
/notes="Organ: Mixed; Vector: pCDNA3.1; Site\_1: EcoRI;  
Site\_2: XhoI; Adult tissue from eye, lung, aorta,  
pulmonary artery and brain; neonatal tissue from liver,  
spleen, thymus, lung, kidney, aorta and brain; 50% is from  
cardiac tissues."

ORIGIN  
Query Match 77.4%; Score 17.8; DB 8; Length 704;  
Best Local Similarity 90.5%; Pred. No. 3.5e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTT 21  
|||||  
DB 136 TGAGAGGCGAGTAGAAGATT 116  
|||||

RESULT 60  
CL583020/c  
LOCUS  
DEFINITION CL583020 OB\_Ba0048023.3 Oryza brachyantha genomic clone  
OB\_Ba0048023 3', genomic survey sequence.  
739 bp DNA linear GSS 10-JUN-2004

ACCESSION CL583020  
VERSION CL583020.1 GI:48541769  
KEYWORDS GSS.  
SOURCE Oryza brachyantha  
ORGANISM Oryza brachyantha  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 739)

REFERENCE Kim, H., Yu, Y., Stum, D., Yost, D., Rao, K., Luo, M., Jetty, R.,  
Kudrna, D., Muller, C., Hatfield, J., Soderlund, C. and Wing, R.  
OMAP Project

ORIGIN  
Query Match 77.4%; Score 17.8; DB 10; Length 739;  
Best Local Similarity 90.5%; Pred. No. 3.6e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTT 21  
|||||  
DB 618 TGAGAGGCGAGTAGAACCAT 598  
|||||

Search completed: January 28, 2006, 01:14:06  
Job time: 1877.03 secs

Unpublished (2004)  
Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 3259  
Email: rwing@genome.arizona.edu  
PCR Primers  
FORWARD: TAA TAC GAC TCA CTA TAG GG  
BACKWARD: CAC TCA TTA GGC ACC CCA  
Insert Length: 145000 Std Error: 0.00  
Plate: 0048 row: 0 column: 23  
Seq primer: CAC TCA TTA GGC ACC CCA  
Class: BAC ends.  
Location/Qualifiers  
1..739  
/organism="Oryza brachyantha"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4533"  
/clone="OB\_Ba0048023"  
/tissue\_type="leaves"  
/dev\_stage="mature"  
/lab\_host="DH10B"  
/clone\_lib="OB\_Ba"  
/note="Vector: pGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

JOURNAL COMMENT

***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:56:54 ; Search time 64.6067 Seconds  
(without alignments)  
632.812 Million cell updates/sec

Title: US-10-716-005-1

Perfect score: 23  
Sequence: 1 tgaagagcagtagaagcttag 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : Issued Patents NA:  
1: /cgn2\_6/prodata/1/ina/1/COMB.seq.\*  
2: /cgn2\_6/prodata/1/ina/5/COMB.seq.\*  
3: /cgn2\_6/prodata/1/ina/6A/COMB.seq.\*  
4: /cgn2\_6/prodata/1/ina/6B/COMB.seq.\*  
5: /cgn2\_6/prodata/1/ina/H/COMB.seq.\*  
6: /cgn2\_6/prodata/1/ina/PCTUS/COMB.seq.\*  
7: /cgn2\_6/prodata/1/ina/PP/COMB.seq.\*  
8: /cgn2\_6/prodata/1/ina/RE/COMB.seq.\*  
9: /cgn2\_6/prodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.8	86.1	1378	US-10-081-923-14	Sequence 14, Appl
2	19.8	86.1	1379	US-10-081-923-7	Sequence 7, Appl
3	19.8	86.1	1379	US-10-081-923-13	Sequence 13, Appl
4	19.8	86.1	1379	US-10-081-923-16	Sequence 16, Appl
5	19.8	86.1	1384	US-10-081-923-8	Sequence 8, Appl
6	19.8	86.1	1384	US-10-081-923-10	Sequence 10, Appl
7	19.8	86.1	1385	US-10-081-923-9	Sequence 9, Appl
8	19.8	86.1	1390	US-10-081-923-11	Sequence 11, Appl
9	19.8	86.1	1393	US-10-081-923-15	Sequence 15, Appl
10	19.8	86.1	1447	US-10-081-923-17	Sequence 17, Appl
11	19.8	86.1	1803	US-10-081-923-5	Sequence 5, Appl
12	18.2	79.1	85122	US-09-949-016-14693	Sequence 14693, A
13	18.2	79.1	119214	US-09-949-016-12507	Sequence 12507, A
14	18.2	79.1	1664976	US-08-916-421B-1	Sequence 1, Appl
15	18.2	79.1	1664976	US-09-692-570-1	Sequence 1, Appl
16	17.8	77.4	222691	US-09-949-016-11762	Sequence 11762, A
17	17.8	77.4	222697	US-09-949-016-115842	Sequence 15842, A
18	17.8	77.4	1664976	US-08-916-421B-1	Sequence 1, Appl
19	17.8	77.4	1664976	US-09-692-570-1	Sequence 1, Appl
20	17.4	75.7	21	US-10-081-923-1	Sequence 1, Appl
21	17.4	75.7	504	US-09-621-976-9879	Sequence 9879, Ap
22	17.2	74.8	601	US-09-949-016-153906	Sequence 153906,
23	17.2	74.8	3929	US-09-799-451-884	Sequence 884, App
24	17.2	74.8	10411	US-08-961-527-89	Sequence 89, Appl

C 25	17.2	74.8	248968	3	US-09-949-016-12614	Sequence 12614, A
C 26	17.2	74.8	250958	3	US-09-949-016-16061	Sequence 16061, A
C 27	16.8	73.0	601	3	US-09-949-016-147767	Sequence 147767,
C 28	16.8	73.0	2061	3	US-09-328-352-177	Sequence 177, App
C 29	16.8	73.0	10421	3	US-09-949-016-16577	Sequence 16577, A
C 30	16.8	73.0	10421	3	US-09-949-016-16578	Sequence 16578, A
C 31	16.8	73.0	44554	3	US-09-949-016-12787	Sequence 12787, A
C 32	16.8	73.0	44555	3	US-09-949-016-12043	Sequence 12043, A
C 33	16.6	72.2	262	3	US-09-513-999C-31242	Sequence 31242, A
C 34	16.6	72.2	461	3	US-09-621-976-3711	Sequence 3711, Ap
C 35	16.6	72.2	601	3	US-09-949-016-195720	Sequence 195720,
C 36	16.6	72.2	852	3	US-09-023-635-881	Sequence 881, App
C 37	16.6	72.2	1907	3	US-08-826-611-1	Sequence 1, Appl
C 38	16.6	72.2	2083	3	US-09-221-017B-938	Sequence 938, App
C 39	16.6	72.2	7785	3	US-09-949-016-12380	Sequence 12380, A
C 40	16.6	72.2	7786	3	US-09-949-016-15590	Sequence 15590, A
C 41	16.6	72.2	41815	3	US-09-949-016-17447	Sequence 17447, A
C 42	16.6	72.2	110402	3	US-09-949-016-17295	Sequence 17295, A
C 43	16.6	72.2	110403	3	US-09-949-016-12741	Sequence 12741, A
C 44	16.6	72.2	115508	3	US-09-949-016-11800	Sequence 11800, A
C 45	16.6	72.2	115508	3	US-09-949-016-14826	Sequence 14826, A
C 46	16.6	72.2	115508	3	US-09-949-016-14827	Sequence 14827, A
C 47	16.4	71.3	62474	3	US-09-949-016-15182	Sequence 15182, A
C 48	16.4	71.3	69709	3	US-09-949-016-15784	Sequence 15784, A
C 49	16.4	71.3	108060	3	US-09-949-016-13159	Sequence 13159, A
C 50	16.4	71.3	121234	3	US-09-949-016-14142	Sequence 14142, A
C 51	16.4	71.3	140315	3	US-09-949-016-14141	Sequence 14141, A
C 52	16.4	71.3	146039	3	US-09-949-016-12449	Sequence 12449, A
C 53	16.4	71.3	678533	3	US-09-949-016-14577	Sequence 14577, A
C 54	16.4	71.3	678533	3	US-09-949-016-14578	Sequence 14578, A
C 55	16.2	70.4	601	3	US-09-949-016-84403	Sequence 84403, A
C 56	16.2	70.4	601	3	US-09-949-016-84404	Sequence 84404, A
C 57	16.2	70.4	601	3	US-09-949-016-165413	Sequence 165413,
C 58	16.2	70.4	601	3	US-09-949-016-165414	Sequence 165414,
C 59	16.2	70.4	601	3	US-09-949-016-165415	Sequence 165415,
C 60	16.2	70.4	601	3	US-09-949-016-165470	Sequence 165470,

ALIGNMENTS

RESULT 1  
US-10-081-923-14  
; Sequence 14, Application US/10081923  
; Patent No. 6593093  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; APPLICANT: Cockerill, Franklin R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/081,923  
; CURRENT FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 1378  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: pta1 sequence from isolate no. 10  
US-10-081-923-14

Query Match 86.1%; Score 19.8; DB 3; Length 1378;  
Best Local Similarity 91.3%; Pred. No. 17;  
Matches 21; Conservativeness 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23

Db 9 TGAATGCGTAGAAGCTTAG 31

RESULT 2

US-10-081-923-7

```
; Sequence 7, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 6
US-10-081-923-7

Query Match      86.1%; Score 19.8; DB 3; Length 1379;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 9 TGAAAATGCAGTAGAAGCTTAG 31

RESULT 3
US-10-081-923-13
; Sequence 13, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 4
US-10-081-923-13

Query Match      86.1%; Score 19.8; DB 3; Length 1379;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 9 TGAAAATGCAGTAGAAGCTTAG 31

RESULT 4
US-10-081-923-16
; Sequence 16, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 8
US-10-081-923-16

Query Match      86.1%; Score 19.8; DB 3; Length 1384;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 9 TGAAAATGCAGTAGAAGCTTAG 31

RESULT 5
US-10-081-923-8
; Sequence 8, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 5
US-10-081-923-8

Query Match      86.1%; Score 19.8; DB 3; Length 1384;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 10 TGAAAATGCAGTAGAAGCTTAG 32

RESULT 6
US-10-081-923-10
; Sequence 10, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptol sequence from isolate no. 8
US-10-081-923-10

Query Match      86.1%; Score 19.8; DB 3; Length 1384;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 10 TGAAAATGCAGTAGAAGCTTAG 32
```



```
RESULT 7
US-10-081-923-9
; Sequence 9, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptai sequence from isolate no. 7
US-10-081-923-9
Query Match      86.1%; Score 19.8; DB 3; Length 1385;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 8
US-10-081-923-11
; Sequence 11, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptai sequence from isolate no. 9
US-10-081-923-11
Query Match      86.1%; Score 19.8; DB 3; Length 1390;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 46

RESULT 9
US-10-081-923-15
; Sequence 15, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
```

```
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptai sequence from isolate no. 3
US-10-081-923-15
Query Match      86.1%; Score 19.8; DB 3; Length 1393;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 10
US-10-081-923-17
; Sequence 17, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptai sequence from isolate no. 11
US-10-081-923-17
Query Match      86.1%; Score 19.8; DB 3; Length 1447;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 10 TGAATAATGCAGTAGAAGCTTAG 32

RESULT 11
US-10-081-923-5
; Sequence 5, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptai sequence from Oklahoma University M1 strain
; PUBLICATION INFORMATION:
; AUTHORS: Ferretti et al.
; JOURNAL: Proc. Natl. Acad. Sci. USA
; VOLUME: 98
; PAGES: 4658-4663
```

```
; DATE: 2001-01-01
US-10-081-923-5

Query Match      86.1%; Score 19.8; DB 3; Length 1803;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCAGAAGGCAGTAGAAGCTTAG 23
Db 177 TGAATAATGCAGTAGAAGCTTAG 199

RESULT 12
US-09-949-016-14693
; Sequence 14693, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14693
; LENGTH: 85122
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(85122)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14693

Query Match      79.1%; Score 18.2; DB 3; Length 85122;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCAGAAGGCAGTAGAAGCTTAG 23
Db 30432 TCAAAAGGCAGTAGAAGCTTAG 30454

RESULT 13
US-09-949-016-12507
; Sequence 12507, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12507
; LENGTH: 119214
; TYPE: DNA
; ORGANISM: Human
```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(119214)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12507

Query Match      79.1%; Score 18.2; DB 3; Length 119214;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
Db 65075 TCAAAAGGCAGTAGAAGCTTAG 65097

RESULT 14
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Jannaschii
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
```

```

; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)

```

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
;
US-08-916-421B-1

```

Query Match 79.1%; Score 18.2; DB 3; Length 1664976;

Best Local Similarity 87.0%; Pred. No. 1.4e+02;

Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 TGAGAGGCAGTAGAAGCTTAG 23
Db 169531 TGAAGGCATTAAAAAGCTTAG 169553

```

# RESULT 15

US-09-692-570-1

; Sequence 1, Application US/09692570

; Patent No. 6797466

; GENERAL INFORMATION:

; APPLICANT: Bult et al.

; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus

; Patent No. 6797466

; TITLE OF INVENTION: jannaachii

; FILE REFERENCE: PB275C1

; CURRENT APPLICATION NUMBER: US/09/692,570

; CURRENT FILING DATE: 2003-01-14

; PRIOR APPLICATION NUMBER: US 60/024,428

; PRIOR FILING DATE: 1996-08-22

; PRIOR APPLICATION NUMBER: US 08/916,421

; PRIOR FILING DATE: 1997-08-22

```
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
```

RESULT 16  
US-09-949-016-11762/c  
; Sequence 11762, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11762  
; LENGTH: 222691  
; TYPE: DNA  
; ORGANISM: Human

```

RESULT 18
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequen
; Patent No. 6503729
; FILE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/034,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature

```

```

; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature

```

```
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600592)..(600592)
```

---

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match      77.4%; Score 17.8; DB 3; Length 1664976;
Best Local Similarity 90.5%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      3 AGAAGCGCAGTAGAAGGCTTAG 23
      ||||| ||||| ||||| |||||
Db      1305980 AGAAGCGCAGATGAAGGCTTAG 1305960

RESULT 19
US-09-692-570-1/c
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
; Patent No. 6797466
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; PRIORITY FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1 :
; LENGTH: 1663976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc feature
```

LOCATION: (622708)..(622708)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (657081)..(657081)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (657203)..(657203)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (674435)..(674435)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (682442)..(682442)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (713652)..(713652)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (741684)..(741684)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (779455)..(779455)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (779676)..(779676)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (855539)..(855539)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (871619)..(871619)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (1084830)..(1084830)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (1096846)..(1096846)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (1119881)..(1119881)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (1130881)..(1130881)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (1310988)..(1310988)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (1313224)..(1313224)	
OTHER INFORMATION: n equals a, t, c, or g	
FEATURE:	
NAME/KEY: misc feature	
LOCATION: (1349473)..(1349473)	
OTHER INFORMATION: n equals a, t, c, or g	

Matches	19;	Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
Qy	3	AGAAGGCAGTAGAAAGCTTAG	23						
Db	1305980	AGAAGGCAGATGAAGCTTAG	1305960						
RESULT 20									
US-10-081-923-1									
; Sequence 1, Application US/10081923									
; Patent No. 6593093									
; GENERAL INFORMATION:									
; APPLICANT: Uhl, James R.									
; APPLICANT: Cockerill, Franklin R.									
; TITLE OF INVENTION: Detection of Group A Streptococcus									
; FILE REFERENCE: 07039-306001									
; CURRENT APPLICATION NUMBER: US/10/081,923									
; CURRENT FILING DATE: 2002-07-02									
; NUMBER OF SEQ ID NOS: 17									
; SOFTWARE: FastSEQ for Windows Version 4.0									
; SEQ ID NO 1									
; LENGTH: 21									
; TYPE: DNA									
; ORGANISM: Artificial Sequence									
; FEATURE:									
; OTHER INFORMATION: Oligonucleotide									
US-10-081-923-1									
Query Match		75.7%	Score 17.4;	DB 3;	Length 21;				
Best Local Similarity		94.7%	Pred. No. 1.4e+00;						
Matches	18;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	5	AAGGCAGTAGAAAGCTTAG	23						
Db	2	AATGCAGTAGAAGCTTAG	20						
RESULT 21									
US-09-621-976-9879									
; Sequence 9879, Application US/09621976									
; Patent No. 6639063									
; GENERAL INFORMATION:									
; APPLICANT: Dumas Milne Edwards, J.B.									
; APPLICANT: Jobert, S.									
; APPLICANT: Giordano, J.V.									
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.									
; FILE REFERENCE: GENSET.054PR2									
; CURRENT APPLICATION NUMBER: US/09/621,976									
; CURRENT FILING DATE: 2000-07-21									
; NUMBER OF SEQ ID NOS: 19335									
; SOFTWARE: Patent.pm									
; SEQ ID NO 9879									
; LENGTH: 504									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
US-09-621-976-9879									
Query Match		75.7%	Score 17.4;	DB 3;	Length 504;				
Best Local Similarity		94.7%	Pred. No. 1.8e+00;						
Matches	18;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Qy	1	TGAGAAGGCAGTAGAAAGC	19						
Db	10	TGAGAAGGCAGCAGAAAGC	28						
RESULT 22									
US-09-949-016-153906/c									
; Sequence 153906, Application US/09949016									
; Patent No. 6812339									
; GENERAL INFORMATION:									
; APPLICANT: VENTER, J. Craig et al.									
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED									



;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF

;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 153906  
;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-153906

Query Match 74.8%; Score 17.2; DB 3; Length 601;  
Best Local Similarity 86.4%; Pred. No. 2.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22  
|||||  
DB 416 TGAGAAGGCAGTAGAAGCTTA 395

RESULT 23  
US-09-799-451-884  
;; Sequence 884, Application US/09799451  
;; Patent No. 6783969  
;; GENERAL INFORMATION:  
;; APPLICANT: Tang, Y. Tom  
;; APPLICANT: Zhou, Ping  
;; APPLICANT: Goodrich, Ryle  
;; APPLICANT: Asundi, Vinod  
;; APPLICANT: Ren, Feiyang  
;; APPLICANT: Zhang, Jie  
;; APPLICANT: Xue, Aigong J.  
;; APPLICANT: Zhao, Qing A.  
;; APPLICANT: Wang, Jian-Rui  
;; APPLICANT: Ma, Yungqing  
;; APPLICANT: Yamazaki, Victoria  
;; APPLICANT: Chen, Rui-hong  
;; APPLICANT: Wang, Zhiwei  
;; APPLICANT: Wang, Dunrui  
;; APPLICANT: Yang, Yonghong  
;; APPLICANT: Wehrman, Tom  
;; APPLICANT: Ghosh, Reena  
;; APPLICANT: Dmanac, Radoje T.  
;; TITLE OF INVENTION: No. 6783969el Nucleic Acids and  
;; TITLE OF INVENTION: Polypeptides  
;; FILE REFERENCE: 803  
;; CURRENT APPLICATION NUMBER: US/09/799,451  
;; CURRENT FILING DATE: 2001-03-05  
;; NUMBER OF SEQ ID NOS: 948  
;; SOFTWARE: pt\_FL\_genes Version 2.0  
;; SEQ ID NO 884  
;; LENGTH: 3929  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (1307)..(1621)  
US-09-799-451-884

Query Match 74.8%; Score 17.2; DB 3; Length 3929;  
Best Local Similarity 86.4%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23  
|||||  
DB 3625 GAGAAGGCAGTAGAAGCTTAG 3646

RESULT 24  
US-08-961-527-89  
;; Sequence 89, Application US/08961527  
;; Patent No. 6420135  
;; GENERAL INFORMATION:  
;; APPLICANT: Charles Kunsch  
;; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
;; NUMBER OF SEQUENCES: 391  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Human Genome Sciences, Inc.  
;; STREET: 9410 Key West Avenue  
;; CITY: Rockville  
;; STATE: Maryland  
;; COUNTRY: USA  
;; ZIP: 20850  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
;; COMPUTER: HP Vectra 486/33  
;; OPERATING SYSTEM: MSDOS version 6.2  
;; SOFTWARE: ASCII Text  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/961,527  
;; FILING DATE:  
;; CLASSIFICATION: 424  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER:  
;; FILING DATE:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Brookes, A. Anders  
;; REGISTRATION NUMBER: 36,373  
;; REFERENCE/DOCKET NUMBER: PB340P1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (301) 309-8504  
;; TELEFAX: (301) 309-8512  
;; INFORMATION FOR SEQ ID NO: 89:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 10411 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
US-08-961-527-89

Query Match 74.8%; Score 17.2; DB 3; Length 10411;  
Best Local Similarity 86.4%; Pred. No. 2.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22  
|||||  
DB 9575 TGAGAAGGCAGTAGAAGCTTA 9596

RESULT 25  
US-09-949-016-12614/c  
;; Sequence 12614, Application US/09949016  
;; Patent No. 6812339  
;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 12614

```
; LENGTH: 248968
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(248968)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12614

Query Match          74.8%; Score 17.2; DB 3; Length 248968;
Best Local Similarity 86.4%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22
   ||||||| |||
DB 116118 TGAGAAGGCAGTAGAAGCTTA 116097

RESULT 26
US-09-949-016-16061/c
; Sequence 16067, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16061
; LENGTH: 250958
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)...(250958)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-16061

Query Match          74.8%; Score 17.2; DB 3; Length 250958;
Best Local Similarity 86.4%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTA 22
   ||||||| |||
DB 166108 TGAGAAGGCAGTAGAAGCTTA 166087

RESULT 27
US-09-949-016-147767
; Sequence 147767, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
```

```
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 147767
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-147767

Query Match          73.0%; Score 16.8; DB 3; Length 601;
Best Local Similarity 81.8%; Pred. No. 3.4e+02;
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAGCTTAG 23
   ||||||| |||
DB 292 GAGAAGGCAGTAGAAGCTTAG 313

RESULT 28
US-09-328-352-177/c
; Sequence 177, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 177
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-177

Query Match          73.0%; Score 16.8; DB 3; Length 2061;
Best Local Similarity 90.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCT 20
   ||||||| |||
DB 522 TGAGAAGTCAGAAGAAAGCT 503

RESULT 29
US-09-949-016-16577
; Sequence 16577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16577
; LENGTH: 10421
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16577

Query Match          73.0%; Score 16.8; DB 3; Length 10421;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      4 GAAGGCAGTAGAAGCTTAG 23
Db      2847 GAAGGCAGTAGGAAGTTTAG 2866

RESULT 30
US-09-949-016-16578
; Sequence 16578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16578
; LENGTH: 10421
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16578

Query Match      73.0%; Score 16.8; DB 3; Length 10421;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 GAAGGCAGTAGAAGCTTAG 23
Db      2847 GAAGGCAGTAGGAAGTTTAG 2866

RESULT 31
US-09-949-016-12787
; Sequence 12787, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12787
; LENGTH: 44554
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12787

Query Match      73.0%; Score 16.8; DB 3; Length 44554;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 GAAGGCAGTAGAAGCTTAG 23
Db      36981 GAAGGCAGTAGGAAGTTTAG 37000

RESULT 32
US-09-949-016-12043
; Sequence 12043, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12043
; LENGTH: 44555
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12043

Query Match      73.0%; Score 16.8; DB 3; Length 44555;
Best Local Similarity 90.0%; Pred. No. 4.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 GAAGGCAGTAGAAGCTTAG 23
Db      36981 GAAGGCAGTAGGAAGTTTAG 37000

RESULT 33
US-09-513-999C-31242
; Sequence 31242, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2,REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 31242
; LENGTH: 262
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 143
; OTHER INFORMATION: w=a or t
US-09-513-999C-31242

Query Match      72.2%; Score 16.6; DB 3; Length 262;
Best Local Similarity 82.6%; Pred. No. 3.8e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TCAGAGGCAGTAGAAGCTTAG 23
Db      164 TCAGAGGCATTAGAAGGCATAG 186

RESULT 34
US-09-621-976-3711
; Sequence 3711, Application US/09621976
```

; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET 054PR2  
; CURRENT APPLICATION NUMBER: US/09/621,976  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 3711  
; LENGTH: 461  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 161..313  
; NAME/KEY: misc\_feature  
; LOCATION: 327  
; OTHER INFORMATION: n=a, g, c or t  
US-09-621-976-3711

Query Match 72.2%; Score 16.6; DB 3; Length 461;  
Best Local Similarity 82.6%; Pred. No. 4e+02; 4; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 4;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
||||| ||||| ||||| |||||  
Db 118 TGAGAATGCAGCAGAGAAGTAG 140

RESULT 35  
US-09-949-016-195720  
; Sequence 195720, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 195720  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-195720

Query Match 72.2%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 4.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
||||| ||||| ||||| |||||  
Db 325 TAAGAAGGCAGCAGACATCTTAG 347

RESULT 36  
US-09-023-655-881/c  
; Sequence 881, Application US/09023655  
; Patent No. 6607879  
; GENERAL INFORMATION:  
; APPLICANT: Cocks, Benjamin G.  
; APPLICANT: Susan G. Stuart

; APPLICANT: Jeffrey J. Seilhamer  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
; TITLE OF INVENTION: EXPRESSION  
; NUMBER OF SEQUENCES: 1508  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/023,655  
; FILING DATE: HERewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, Karen J.  
; REGISTRATION NUMBER: 37,071  
; REFERENCE/DOCKET NUMBER: PA-0001 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650) 855-0555  
; TELEFAX: (650) 845-4166  
; INFORMATION FOR SEQ ID NO: 881:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 852 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GENBANK  
; CLONE: g1122218  
US-09-023-655-881

Query Match 72.2%; Score 16.6; DB 3; Length 852;  
Best Local Similarity 82.6%; Pred. No. 4.2e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAGAAGGCAGTAGAAGCTTAG 23  
||||| ||||| ||||| |||||  
Db 828 TGGGAAGGCAGCAGATGCTTAG 806

RESULT 37  
US-08-826-611-1  
; Sequence 1, Application US/08826611  
; Patent No. 6031154  
; GENERAL INFORMATION:  
; APPLICANT: Bennett, Alan B.  
; APPLICANT: Kanayama, Yoshinori  
; TITLE OF INVENTION: Fructokinase Genes and Their Use in  
; TITLE OF INVENTION: Metabolic Engineering of Fruit Sweetness  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

	US-08-826-611-1	Query Match Best Local Similarity 72.2%; Score 16.6; DB 3; Length 1907; Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 TGAGAGGCGTAGAAGCTTAC 23	
Db	1312 TGAGAGGCGAGCAACAAGTTTG 1334	
RESULT 38		
US-09-221-017B-938/c	Sequence 938, Application US/09221017B	
Patent No. 6444799	GENERAL INFORMATION:	
APPLICANT: ROSS, BRUCE C.	TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF	
NUMBER OF SEQUENCES: 1120	CORRESPONDENCE ADDRESS:	
ADDRESS: MORRISON & FOERSTER	STREET: 755 PAGE MILL ROAD	
CITY: Palo Alto	STATE: CA	
COUNTRY: USA	ZIP: 94304-1018	
COMPUTER READABLE FORM:	MEDIUM TYPE: Diskette	
OPERATING SYSTEM: Windows	SOFTWARE: FastSeq for Windows Version 2.0b	
PRIOR APPLICATION DATA:	FILING DATE: 23-DEC-1998	
CLASSIFICATION:	PRIOR APPLICATION NUMBER: PP1546	
APPLICATION DATA:	FILING DATE: 30-JAN-1998	
APPLICATION NUMBER: PP2911	FILING DATE: 09-APR-1998	
PRIOR APPLICATION DATA:	APPLICATION NUMBER: PCT/AU98/01023	
FILING DATE: 10-DEC-1998	ATTORNEY/AGENT INFORMATION:	
NAME: Monroy, Gladys H	REGISTRATION NUMBER: 32,430	
REFERENCE/DOCKET NUMBER: 27340-20021.00		

	US-08-826-611-1	Query Match Best Local Similarity 72.2%; Score 16.6; DB 3; Length 1907; Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 TGAGAGGCGTAGAAGCTTAC 23	
Db	1312 TGAGAGGCGAGCAACAAGTTTG 1334	
RESULT 38		
US-09-221-017B-938/c	Sequence 938, Application US/09221017B	
Patent No. 6444799	GENERAL INFORMATION:	
APPLICANT: ROSS, BRUCE C.	TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF	
NUMBER OF SEQUENCES: 1120	CORRESPONDENCE ADDRESS:	
ADDRESS: MORRISON & FOERSTER	STREET: 755 PAGE MILL ROAD	
CITY: Palo Alto	STATE: CA	
COUNTRY: USA	ZIP: 94304-1018	
COMPUTER READABLE FORM:	MEDIUM TYPE: Diskette	
OPERATING SYSTEM: Windows	SOFTWARE: FastSeq for Windows Version 2.0b	
PRIOR APPLICATION DATA:	FILING DATE: 23-DEC-1998	
CLASSIFICATION:	PRIOR APPLICATION NUMBER: PP1546	
APPLICATION DATA:	FILING DATE: 30-JAN-1998	
APPLICATION NUMBER: PP2911	FILING DATE: 09-APR-1998	
PRIOR APPLICATION DATA:	APPLICATION NUMBER: PCT/AU98/01023	
FILING DATE: 10-DEC-1998	ATTORNEY/AGENT INFORMATION:	
NAME: Monroy, Gladys H	REGISTRATION NUMBER: 32,430	
REFERENCE/DOCKET NUMBER: 27340-20021.00		

	US-08-826-611-1	Query Match Best Local Similarity 72.2%; Score 16.6; DB 3; Length 1907; Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 TGAGAGGCGTAGAAGCTTAC 23	
Db	1312 TGAGAGGCGAGCAACAAGTTTG 1334	
RESULT 38		
US-09-221-017B-938/c	Sequence 938, Application US/09221017B	
Patent No. 6444799	GENERAL INFORMATION:	
APPLICANT: ROSS, BRUCE C.	TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF	
NUMBER OF SEQUENCES: 1120	CORRESPONDENCE ADDRESS:	
ADDRESS: MORRISON & FOERSTER	STREET: 755 PAGE MILL ROAD	
CITY: Palo Alto	STATE: CA	
COUNTRY: USA	ZIP: 94304-1018	
COMPUTER READABLE FORM:	MEDIUM TYPE: Diskette	
OPERATING SYSTEM: Windows	SOFTWARE: FastSeq for Windows Version 2.0b	
PRIOR APPLICATION DATA:	FILING DATE: 23-DEC-1998	
CLASSIFICATION:	PRIOR APPLICATION NUMBER: PP1546	
APPLICATION DATA:	FILING DATE: 30-JAN-1998	
APPLICATION NUMBER: PP2911	FILING DATE: 09-APR-1998	
PRIOR APPLICATION DATA:	APPLICATION NUMBER: PCT/AU98/01023	
FILING DATE: 10-DEC-1998	ATTORNEY/AGENT INFORMATION:	
NAME: Monroy, Gladys H	REGISTRATION NUMBER: 32,430	
REFERENCE/DOCKET NUMBER: 27340-20021.00		

	US-08-826-611-1	Query Match Best Local Similarity 72.2%; Score 16.6; DB 3; Length 1907; Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 TGAGAGGCGTAGAAGCTTAC 23	
Db	1312 TGAGAGGCGAGCAACAAGTTTG 1334	
RESULT 38		
US-09-221-017B-938/c	Sequence 938, Application US/09221017B	
Patent No. 6444799	GENERAL INFORMATION:	
APPLICANT: ROSS, BRUCE C.	TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF	
NUMBER OF SEQUENCES: 1120	CORRESPONDENCE ADDRESS:	
ADDRESS: MORRISON & FOERSTER	STREET: 755 PAGE MILL ROAD	
CITY: Palo Alto	STATE: CA	
COUNTRY: USA	ZIP: 94304-1018	
COMPUTER READABLE FORM:	MEDIUM TYPE: Diskette	
OPERATING SYSTEM: Windows	SOFTWARE: FastSeq for Windows Version 2.0b	
PRIOR APPLICATION DATA:	FILING DATE: 23-DEC-1998	
CLASSIFICATION:	PRIOR APPLICATION NUMBER: PP1546	
APPLICATION DATA:	FILING DATE: 30-JAN-1998	
APPLICATION NUMBER: PP2911	FILING DATE: 09-APR-1998	
PRIOR APPLICATION DATA:	APPLICATION NUMBER: PCT/AU98/01023	
FILING DATE: 10-DEC-1998	ATTORNEY/AGENT INFORMATION:	
NAME: Monroy, Gladys H	REGISTRATION NUMBER: 32,430	
REFERENCE/DOCKET NUMBER: 27340-20021.00		

	US-08-826-611-1	Query Match Best Local Similarity 72.2%; Score 16.6; DB 3; Length 1907; Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy	1 TGAGAGGCGTAGAAGCTTAC 23	
Db	1312 TGAGAGGCGAGCAACAAGTTTG 1334	
RESULT 38		
US-09-221-017B-938/c	Sequence 938, Application US/09221017B	
Patent No. 6444799	GENERAL INFORMATION:	
APPLICANT: ROSS, BRUCE C.	TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF	
NUMBER OF SEQUENCES: 1120	CORRESPONDENCE ADDRESS:	
ADDRESS: MORRISON & FOERSTER	STREET: 755 PAGE MILL ROAD	
CITY: Palo Alto	STATE: CA	
COUNTRY: USA	ZIP: 94304-1018	
COMPUTER READABLE FORM:	MEDIUM TYPE: Diskette	
OPERATING SYSTEM: Windows	SOFTWARE: FastSeq for Windows Version 2.0b	
PRIOR APPLICATION DATA:	FILING DATE: 23-DEC-1998	
CLASSIFICATION:	PRIOR APPLICATION NUMBER: PP1546	
APPLICATION DATA:	FILING DATE: 30-JAN-1998	
APPLICATION NUMBER: PP2911	FILING DATE: 09-APR-1998	
PRIOR APPLICATION DATA:	APPLICATION NUMBER: PCT/AU98/01023	
FILING DATE: 10-DEC-1998	ATTORNEY/AGENT INFORMATION:	
NAME: Monroy, Gladys H	REGISTRATION NUMBER: 32,430	
REFERENCE/DOCKET NUMBER: 27340-20021.00		

	US-08-826-611-1	Query Match Best Local Similarity 72.2%; Score 16.6; DB 3; Length 1907; 
--	-----------------	--

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15590  
; LENGTH: 7786  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15590

Query Match 72.2%; Score 16.6; DB 3; Length 7786;  
Best Local Similarity 82.6%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 4398 TGAGAAGCCTGTAGAAGTTCAG 4420

## RESULT 41

US-09-949-016-17447  
; Sequence 17447, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17447  
; LENGTH: 41815  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17447

Query Match 72.2%; Score 16.6; DB 3; Length 41815;  
Best Local Similarity 82.6%; Pred. No. 5.7e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 11357 TGAGAAGCCTGTAGAAGTTAAG 11379

## RESULT 42

US-09-949-016-17295/c  
; Sequence 17295, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17295  
; LENGTH: 110402  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-17295

Query Match 72.2%; Score 16.6; DB 3; Length 110402;  
Best Local Similarity 82.6%; Pred. No. 6e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 17415 TAAGAAGGCAGCAGACATCTTAG 17393

## RESULT 43

US-09-949-016-12741/c  
; Sequence 12741, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12741  
; LENGTH: 110403  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12741

Query Match 72.2%; Score 16.6; DB 3; Length 110403;  
Best Local Similarity 82.6%; Pred. No. 6e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
Db 17422 TAAGAAGGCAGCAGACATCTTAG 17400

## RESULT 44

US-09-949-016-11800/c  
; Sequence 11800, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11800

```
; LENGTH: 115508
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11800

Query Match          72.2%; Score 16.6; DB 3; Length 115508;
Best Local Similarity 82.6%; Pred. No. 6.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGGCTTAG 23
    |||||
Db 1160 TGAGAGGCGAGGAGGAGGCTCAG 1138

RESULT 47
US-09-949-016-15182
; Sequence 15182, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15182
; LENGTH: 62474
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-15182

Query Match          71.3%; Score 16.4; DB 3; Length 62474;
Best Local Similarity 94.4%; Pred. No. 7.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTGAAGG 19
    |||||
Db 3601 GAGAGGCGAGGAGGAGG 3618

RESULT 48
US-09-949-016-15784/c
; Sequence 15784, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15784
; LENGTH: 69709
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(69709)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15784

Query Match          71.3%; Score 16.4; DB 3; Length 69709;
```

```
; LENGTH: 115508
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-11800

Query Match          72.2%; Score 16.6; DB 3; Length 115508;
Best Local Similarity 82.6%; Pred. No. 6.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGGCTTAG 23
    |||||
Db 1160 TGAGAGGCGAGGAGGAGGCTCAG 1138

RESULT 45
US-09-949-016-14826/c
; Sequence 14826, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14826
; LENGTH: 115508
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14826

Query Match          72.2%; Score 16.6; DB 3; Length 115508;
Best Local Similarity 82.6%; Pred. No. 6.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTGAAGGCTTAG 23
    |||||
Db 1160 TGAGAGGCGAGGAGGAGGCTCAG 1138

RESULT 46
US-09-949-016-14827/c
; Sequence 14827, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14827
; LENGTH: 115508
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14827
```

Best Local Similarity 94.4%; Pred. No. 7.2e+02; DB 3; Length 121234;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||  
Db 14376 GAGAAGGCAGTTGAAAGC 14359

RESULT 49  
US-09-949-016-13159/c  
; Sequence 13159, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13159  
; LENGTH: 108060  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13159

Query Match 71.3%; Score 16.4; DB 3; Length 108060;  
Best Local Similarity 94.4%; Pred. No. 7.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||  
Db 104460 GAGAAGGCAGAAAGC 104443

RESULT 50  
US-09-949-016-14142/c  
; Sequence 14142, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14142  
; LENGTH: 121234  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14142

Query Match 71.3%; Score 16.4; DB 3; Length 121234;  
Best Local Similarity 94.4%; Pred. No. 7.4e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||

Db 117634 GAGAAGGCAGAGAAAGC 117617

RESULT 51  
US-09-949-016-14141/c  
; Sequence 14141, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14141  
; LENGTH: 140315  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14141

Query Match 71.3%; Score 16.4; DB 3; Length 140315;  
Best Local Similarity 94.4%; Pred. No. 7.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||  
Db 136715 GAGAAGGCAGAGAAAGC 136698

RESULT 52  
US-09-949-016-12449/c  
; Sequence 12449, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12449  
; LENGTH: 146039  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(146039)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12449

Query Match 71.3%; Score 16.4; DB 3; Length 146039;  
Best Local Similarity 94.4%; Pred. No. 7.5e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GAGAAGGCAGTAGAAAGC 19  
|||||  
Db 42706 GAGAAGGCAGTTGAAAGC 42689



```
RESULT 53
US-09-949-016-14577/c
; Sequence 14577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14577
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14577

Query Match 71.3%; Score 16.4; DB 3; Length 678533;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAGAGGCGAGTAGAAGC 19
Db 174425 GAGAGGCGAGTAGAAGC 174408

RESULT 54
US-09-949-016-14578/c
; Sequence 14578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14578
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14578

Query Match 71.3%; Score 16.4; DB 3; Length 678533;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAGAGGCGAGTAGAAGC 19
Db 174425 GAGAGGCGAGTAGAAGC 174408

RESULT 55
US-09-949-016-84403
; Sequence 84403, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84403
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-84403

Query Match 70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAGAGGCGAGTAGAAGCTTA 22
Db 145 GAGAGGCGAGTAGAAGCTTA 165

RESULT 56
US-09-949-016-84404
; Sequence 84404, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 84404
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-84404

Query Match 70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 GAGAGGCGAGTAGAAGCTTA 22
Db 66 GAGAGGCGAGTAGAAGCTTA 86

RESULT 57
US-09-949-016-14579/c
; Sequence 14579, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14579
; LENGTH: 678533
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(678533)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-14579

Query Match 71.3%; Score 16.4; DB 3; Length 678533;
Best Local Similarity 94.4%; Pred. No. 7.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GAGAGGCGAGTAGAAGC 19
Db 174425 GAGAGGCGAGTAGAAGC 174408
```

```
RESULT 57
US-09-949-016-165413
; Sequence 165413, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165413
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-165413

Query Match          70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAAGGCAGTAGAAAAGCTTA 22
Db      14 GAGAAGGCAGTAGAAAATTTA 34

RESULT 58
US-09-949-016-165414
; Sequence 165414, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165414
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-165414

Query Match          70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAAGGCAGTAGAAAAGCTTA 22
Db      14 GAGAAGGCAGTAGAAAATTTA 34

RESULT 59
US-09-949-016-165415
; Sequence 165415, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 165415
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-165415

Query Match          70.4%; Score 16.2; DB 3; Length 601;
Best Local Similarity 85.7%; Pred. No. 6.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GAGAAGGCAGTAGAAAAGCTTA 22
Db      14 GAGAAGGCAGTAGAAAATTTA 34

Search completed: January 28, 2006, 01:18:29
Job time : 81.6067 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:49:12 ; Search time 1864.03 Seconds  
(without alignments)  
577.298 Million cell updates/sec

Title: US-10-716-005-1  
Perfect score: 23  
Sequence: 1 tgaagaggcagtagaagcttag 23

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gss1:\*  
10: gb\_gss2:\*  
11: gb\_gss3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	18.8	81.7	468	9 AZ279226	AZ279226 RPCI-23-1
C 2	18.8	81.7	636	9 AZ642726	AZ642726 IM0505G22
C 3	18.8	81.7	669	3 BP955145	BP955145 BP955145
C 4	18.8	81.7	680	11 CR892303	CR892303 Sus scrofa
C 5	18.8	81.7	999	10 CL032011	CL032011 CH216-33M
C 6	18.4	80.0	542	10 CE365658	CE365658 tigr-gss-
C 7	18.4	80.0	628	6 CD695839	CD695839 EST12362
C 8	18.4	80.0	656	3 BI508824	BI508824 BB170010B
C 9	18.4	80.0	709	10 CE409091	CE409091 tigr-gss-
C 10	18.4	80.0	815	10 AG551894	AG551894 Mus muscu
C 11	18.2	79.1	357	10 BB873514	BB873514 BB873514
C 12	18.2	79.1	377	10 CG682385	CG682385 ZMMB0014
C 13	18.2	79.1	429	7 CF982181	CF982181 maj179A02
C 14	18.2	79.1	433	7 CK100703	CK100703 C067P49.5
C 15	18.2	79.1	531	9 BH025854	BH025854 RPCI-24-2
C 16	18.2	79.1	544	5 BU814643	BU814643 N031C09 P
C 17	18.2	79.1	567	6 CA929386	CA929386 MTU2CA.P1
C 18	18.2	79.1	571	6 CF105022	CF105022 mai93C06
C 19	18.2	79.1	601	3 BQ322034	BQ322034 PM2-CT026
C 20	18.2	79.1	615	1 AV270535	AV270535 AV270535
C 21	18.2	79.1	627	9 CE110333	CE110333 tigr-gss-
C 22	18.2	79.1	640	10 CE426147	CE426147 tigr-gss-

ALIGNMENTS

RESULT 1  
AZ279226/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AZ279226 468 bp DNA linear GSS 26-JUL-2000  
RPCI-23-135F5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-135F5,  
genomic survey sequence.

AZ279226 GI:9498128  
GSS.  
Mus musculus (house mouse)

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 468)  
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,  
Akinret,B., Levine,M., McGann,S., Tsengaye,G., Geer,K., Krol,M., de  
Jong,P. and Fraser,C.M.

Mus musculus  
Jong,P. and Fraser,C.M.

Unpublished (1999)

Contact: Shaying Zhao  
Department of Eukaryotic Genomes  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: szhao@tigr.org  
Clones are derived from the mouse BAC library RPCI-23. For BAC  
library availability, please contact Pieter de Jong  
(pieter@jeong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

BZ984463 PUF1180TD  
CX179363 D07\_45-16  
CT008091 Equus cab  
BU946020 AGENCOURT  
B1737502 603358184  
CO431201 UI-M-HX0-  
BU466365 603268708  
CA494633 AGENCOURT  
CZ717363 OC\_Ba004  
DT058135 AGENCOURT  
BC037104 Mus muscu  
BC048862 Mus muscu  
BC038223 Homo sapi  
BH868412 HK25C06.x  
CO747995 ShESTbaa6  
CV629435 M3frt3101  
CV630812 M3frt3105  
CV657147 M3frt3109  
BE142442 RC2-HT014  
BM304832 ShEST4a32  
BF323664 ShEST4a18  
BM305256 ShEST4a38  
AZ851239 2M0153G22  
AQ437381 HS\_51222\_A  
BM303377 ShEST4a58  
AQ595160 HS\_5424\_B  
CD313569 StrPu621.  
CR221772 Reverse e  
CN496517 Mafw2022g  
CB638470 OSUNE006M  
CD278427 T143C0237  
CD219734 CCC1\_58\_H  
CX000060 1v54H07.b  
AZ649204 1M0518J21  
CE179177 tigr-gss-  
AQ541187 RPCI-11-3  
DR108351 JHU162A06  
CL583020 OB\_Ba004

or from Reese ch Genetics (info@resgen.com). BAC end page:  
 http://www.tigr.org/tdb/bac ends/mouse/bac\_end\_intro.html  
 Plate: 135 row: F column: 5  
 Seq primer: SP6  
 Class: BAC ends.

#### FEATURES

source  
 1. .468  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-23-135F5"  
 /sex="Female"  
 /lab\_host="DH10B"  
 /clone\_lib="RPCI-23"  
 /notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:  
 EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methyase. Size  
 selected DNA was cloned into the pBACe3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."

#### ORIGIN

Query Match 81.7%; Score 18.8; DB 9; Length 468;  
 Best Local Similarity 90.9%; Pred. No. 1.2e+03;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TGAGAAGGCAGTAGAAGCTTA 22  
 ||||| ||||| ||||| |||||  
 Db 249 TGAGAAGGCAGTAGAAGCTTA 228

#### RESULT 2

AZ642726/c  
 LOCUS 636 bp DNA linear GSS 14-DBC-2000  
 DEFINITION 1M0505G22R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
 clone UUGC1M0505G22 R, genomic survey sequence.

ACCESSION AZ642726  
 VERSION GSS.  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Mus musculus  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.  
 1 (bases 1 to 636)

REFERENCE  
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
 Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
 Niederhausern,A. and Wright,D.,Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts

TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA

FEATURES  
 source  
 1. .636  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 High quality sequence stop: 636.

#### FEATURES

source  
 1. .468  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"

/clone="UUGC1M0505G22"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi|4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

#### ORIGIN

Query Match 81.7%; Score 18.8; DB 9; Length 636;  
 Best Local Similarity 90.9%; Pred. No. 1.2e+03;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TGAGAAGGCAGTAGAAGCTTA 22  
 ||||| ||||| ||||| |||||  
 Db 37 TGAGAAGGCAGTAGAAGCTTA 16

#### RESULT 3

BP955145  
 LOCUS 669 bp mRNA linear EST 10-AUG-2005  
 DEFINITION BP955145 Euphorbia tirucalli callus induced from stem Euphorbia  
 tirucalli cDNA clone ETC041C09 5', mRNA sequence.

ACCESSION BP955145  
 VERSION BP955145.1 GI:58204501  
 KEYWORDS EST.  
 SOURCE Euphorbia tirucalli  
 ORGANISM Euphorbia tirucalli  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Malpighiales; Euphorbiales; Euphorbiaceae;  
 Euphorbiae; Euphorbia.  
 1 (bases 1 to 669)

REFERENCE  
 AUTHORS Kajikawa,M., Yamato,K.T., Kohzu,Y., Sakata,R., Fukuzawa,H.,  
 Uchida,H. and Ohyama,K.  
 Expressed sequence tags from callus of Euphorbia tirucalli: A  
 resource for genes involved in triterpenoid and sterol biosynthesis  
 Plant Biotechnol. 21, 349-353 (2004)  
 CONTACT: Katsuyuki T Yamato  
 Graduate School of Biostudies, Division of Integrated Life Science  
 Kyoto University  
 Sakyo-ku, Kitashirakawa-iwake-cho, Kyoto, Kyoto 606-8502, Japan  
 Tel: 81-75-753-6453  
 Fax: 81-75-753-6127  
 Email: kyamato@lif.kyoto-u.ac.jp.

FEATURES  
 source  
 1. .669  
 Location/Qualifiers  
 /organism="Euphorbia tirucalli"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:142860"  
 /clone="ETC041C09"  
 /tissue\_type="callus induced from stem"  
 /clone\_lib="Euphorbia tirucalli callus induced from stem"

#### ORIGIN

Query Match 81.7%; Score 18.8; DB 3; Length 669;  
 Best Local Similarity 90.9%; Pred. No. 1.2e+03;

```

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTA 22
Db 448 TGAGAGGCAGCAGCAAAAGTTTA 469

RESULT 4
CR8922303/c
LOCUS
DEFINITION
SUB scrofa BES, genomic survey sequence.
ACCESSION
CR892303
VERSION
CR892303.1 GI:55970967
KEYWORDS
GSS: Bac-end sequence BES; Genome Survey Sequence.
SOURCE
Sus scrofa (pig)
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
REFERENCE
1 (bases 1 to 680)
Rogel-Gaillard,C., Bourgeaux,N., Billault,A., Vaiman,M. and
Chardon,P.
Construction of a swine BAC library: application to the
characterization and mapping of porcine type C endoviral elements
Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)
PUBMED
10449899
REFERENCE
2 (bases 1 to 680)
Chardon,P., Iannuccelli,N., Roig,A., Dossat,C., Demars,J.,
Rogel-Gaillard,C., Roy,A., Schibler,L. and Milan,D.
A physical map of the swine genome
Unpublished
3 (bases 1 to 680)
Genoscope.
Direct Submission
Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
FEATURES
source
1..680
/organism="Sus scrofa"
/mol_type="genomic DNA"
/strain="Large White"
/db_xref="taxon:9823"
/clone="BT0179H06"
/sex="male"
/cell_type="fibroblast"
/clone_lib="S8AB"
/note="Genoscope sequence ID : IH0AAA17AD12RM1"
ORIGIN

Query Match 81.7%; Score 18.8; DB 11; Length 680;
Best Local Similarity 90.9%; Pred. No. 1.2e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GAGCAGGCAGTAGAAGCTTAG 23
Db 551 GAGCAGGCAGTAGAAGCTTGG 530

RESULT 5
CL032011/c
LOCUS
DEFINITION
CH216-33M22_sp6.1 CH216 Xenopus tropicalis genomic clone
CH216-33M22, genomic survey sequence.
ACCESSION
CL032011
VERSION
CL032011.1 GI:40481281
KEYWORDS
GSS.
Xenopus tropicalis (western clawed frog)
ORGANISM
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 999)
REFERENCE

AUTHORS
Krenitzki,C., Carter,J., McPherson,J., Warren,W., Graves,T.,
Mardis,E. and Wilson,R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: Sp6 ATTTAGGTGACACTATAG
Class: BAC ends
High quality sequence start: 33
High quality sequence stop: 526.
Location/Qualifiers
1..999
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-33M22"
/sex="male"
/cell_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"
ORIGIN

Query Match 81.7%; Score 18.8; DB 10; Length 999;
Best Local Similarity 90.9%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTA 22
Db 898 TGAGAGGTAGAGAAAGCTTA 877

RESULT 6
CE365658
LOCUS
DEFINITION
tigr-g88-dog-17000362151284 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION
CE365658
VERSION
CE365658.1 GI:36583364
KEYWORDS
GSS.
SOURCE
Canis familiaris (dog)
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
REFERENCE
1 (bases 1 to 542)
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
PUBMED
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1..542
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"
FEATURES
source

```

```

ORIGIN
Query Match      80.0%; Score 18.4; DB 10; Length 542;
Best Local Similarity 95.0%; Pred. No. 1.9e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GAAGGCGAGTAGAAGCTTAG 23
    |||||
Db 311 GAAGGCGAGTAGAAGCTCAG 330

RESULT 7
CD695839
LOCUS      628 bp mRNA linear EST 25-JUN-2003
DEFINITION human nasopharynx Homo sapiens cDNA, mRNA sequence.
ACCESSION CD695839
VERSION    CD695839.1 GI:32221773
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominoidea; Homo.
REFERENCE  1 (bases 1 to 628)
AUTHORS   Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and
            Zeng,Y.-X.
TITLE     Transcriptional Gene Expression Profile of Human Nasopharynx
JOURNAL   Unpublished (2003)
COMMENT   Contact: Yixin Zeng
            Cancer Center
            Sun Yat-sen University
            651 Dongfeng Road East, Guangzhou 510060, China
            Tel: 86-1380-9770-743
            Fax: 86-20-8775-4506
            Email: yxzeng@zsusm.edu.cn.
FEATURES   Location/Qualifiers
            source          1..628
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /tissue_type="normal nasopharynx"
                        /clone_lib="human nasopharynx"
                        /notes="ESTs generated from a normal nasopharynx cDNA
                        library from southern Chinese"

ORIGIN
Query Match      80.0%; Score 18.4; DB 6; Length 628;
Best Local Similarity 95.0%; Pred. No. 1.9e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAAGGCGAGTAGAAGCT 20
    |||||
Db 383 TAAGAAGGCGAGTAGAAGCT 402

RESULT 8
BI508824
LOCUS      656 bp mRNA linear EST 08-APR-2002
DEFINITION BBI70010B20G05.5 Bee Brain Normalized/Subtracted Library, BBI7 Apis
            mellifera cDNA clone BBI70010B20G05 5', mRNA sequence.
ACCESSION BI508824
VERSION    BI508824
KEYWORDS   EST.
SOURCE     BBI508824.1 GI:15359198
            Apis mellifera (honey bee)
ORGANISM   Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apoidea; Apis.
            1 (bases 1 to 656)
REFERENCE  1 (bases 1 to 656)
AUTHORS   Whitfield,C.W., Band,M.R., Bonaldo,M.F., Kumar,C.G., Liu,L.,
            Pardinas,J., Robertson,H.M., Soares,B. and Robinson,G.B.
TITLE     Annotated expressed sequence tags and cDNA microarrays for studies
            of brain and behavior in the honey bee

JOURNAL PUBLISHED
COMMENT
Genome Res. 12 (4), 555-566 (2002)
11932240
Contact: Gene E. Robinson
Department of Entomology
University of Illinois
505 S. Goodwin Ave., Urbana, IL 61801, USA
Tel: 217 265 0309
Fax: 217 244 3499
Email: generobi@life.uiuc.edu
This research was funded by the University of Illinois Critical
Research Initiatives Fund and a Burroughs-Wellcome Trust Innovation
Award in Functional Genomics to G.E. Robinson and an NSF
Postdoctoral Fellowship in Bioinformatics to C.W. Whitfield.
PCR PRIMERS
FORWARD: TAATACGACTCACTATAGG
BACKWARD: ATTAACCTCACTAAAG
Plate: BBI70010B20 row: G column: 05
Seq primer: AGCGGATAACAATTTTCACACAGGA
High quality sequence stop: 656.
FEATURES   Location/Qualifiers
            source          1..656
                        /organism="Apis mellifera"
                        /mol_type="mRNA"
                        /strain="mixed strains of European bees, predominantly
                        A.m. ligustica"
                        /db_xref="taxon:7460"
                        /clone="BBI70010B20G05"
                        /sex="female"
                        /tissue_type="brain"
                        /dev_stage="adult worker honey bee"
                        /lab_host="DH10B"
                        /clone_lib="Bee Brain Normalized/Subtracted Library, BBI7"
                        /note="Organ: brain; Vector: pMT3-Pac; Site 1: EcoRI;
                        Site 2: NotI; This BBI7 cDNA library was generated by
                        subtraction of the BBI6 library with 4000 previously
                        sequenced clones. The BBI6 library was contributed by the
                        Soares laboratory and it was constructed and normalized
                        as described by Bonaldo, M.F., Lennon, G. and Soares,
                        M.B. (1996), Genome Research 6(9): 791-806. RNA was
                        prepared from dissected brains of adult worker bees of
                        various ages and various behavioral groups. "
            Query Match      80.0%; Score 18.4; DB 3; Length 656;
            Best Local Similarity 95.0%; Pred. No. 1.9e+03;
            Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGAAGGCGAGTAGAAGCT 20
    |||||
Db 116 TCAGAAGGCGAGTAGAAGCT 135

RESULT 9
CE409091/c
LOCUS      709 bp DNA linear GSS 27-SEP-2003
DEFINITION tigr-gss-dog-17000362259182 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION CE409091
VERSION    CE409091.1 GI:36663905
KEYWORDS   GSS.
SOURCE     Canis familiaris (dog)
ORGANISM   Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
            Canis.
            1 (bases 1 to 709)
REFERENCE  1 (bases 1 to 709)
AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE     The dog genome: survey sequencing and comparative analysis
JOURNAL   Science 301 (5641), 1898-1903 (2003)
PUBLISHED 14512627
COMMENT   Contact: Kirkness EF

```

The Institute for Genomic Research  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-0200  
Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

# FEATURES

## source

Location/Qualifiers  
1..709  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="dog Library"  
/notes="Site 1: BstXI; Libraries were prepared from  
peripheral blood"

# ORIGIN

Query Match 80.0%; Score 18.4; DB 10; Length 709;  
Best Local Similarity 95.0%; Pred. No. 1.9e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GAAGGCAGTAGAAGCTTAG 23  
|||||  
Db 259 GAAGGCAGTAGAAGCTCAG 240  
|||||

RESULT 10  
AG551894/c  
LOCUS  
DEFINITION  
AG551894 815 bp DNA linear GSS 23-DEC-2004  
Mus musculus molossinus DNA, clone:MSMg01-468G16.TJ, genomic survey  
sequence.  
ACCESSION  
AG551894  
VERSION  
AG551894.1 GI:48312585  
KEYWORDS  
GSS.  
SOURCE  
Mus musculus molossinus (Japanese wild mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
1 Abe, K., Noguchi, H., Tagawa, K., Yuzuriha, M., Toyoda, A., Kojima, T.,  
Ezawa, K., Saitou, N., Hattori, M., Sakaki, Y., Moriwa, K. and  
Shiroishi, T.  
Contribution of Asian mouse subspecies Mus musculus molossinus to  
genomic constitution of strain C57BL/6J, as defined by BAC-end  
sequence-SNP analysis  
Genome Res. 14 (12), 2439-2447 (2004)

JOURNAL  
PUBMED  
15574823  
REFERENCE  
2 (bases 1 to 815)  
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.  
Direct Submission  
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
(E-mail: hattori@gsc.riken.jp, URL: http://hgp.gsc.riken.go.jp/,  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT  
Clones are derived from the mouse BAC library MSMg01. For BAC  
library availability, please contact Kuniya Abe (abe@rtc.riken.jp).  
Tsukuba Institute, Bio Resource Center.  
The Institute of Physical and Chemical Research (RIKEN) 3-1-1  
Koyadai, Tsukuba, 305-0074 Japan  
phone: 81-298-36-9189, fax: 81-298-36-9199  
e-mail: abe@rtc.riken.jp

## PRIMERS

Sequencing : TJ  
LIBRARY  
Vector : pBACe3.6  
R.Site 1 : EcoRI  
R.Site 2 : EcoRI.

# FEATURES

1..815  
/organism="Mus musculus molossinus"

/mol\_type="genomic DNA"  
/sub\_species="molossinus"  
/db\_xref="taxon:57486"  
/clone="MSMg01-468G16.TJ"  
/sex="male"  
/tissue\_type="mixture of kidney and spleen"  
/clone\_lib="MSMg01 Mouse Male BAC Library"

# ORIGIN

Query Match 80.0%; Score 18.4; DB 10; Length 815;  
Best Local Similarity 95.0%; Pred. No. 1.9e+03;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 GAGAGGCAGTAGAAGCTT 21  
|||||  
Db 757 GAGAGGCAGTAGAAGCTT 738  
|||||

RESULT 11  
BB873514/c  
LOCUS  
DEFINITION

BB873514 357 bp mRNA linear EST 27-NOV-2001  
BB873514 RIKEN full-length enriched, adult male intestinal mucosa  
Mus musculus cDNA clone G630046N07 5', mRNA sequence.

ACCESSION  
BB873514  
VERSION  
BB873514.1 GI:17119724  
KEYWORDS  
EST.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muroidea; Muridae; Murinae; Mus.

REFERENCE  
1 (bases 1 to 357)

Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Imotani, K.,  
Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,  
Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,  
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Tanaka, T., Tomaru, A., Toya, T., Watahiki, A., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.  
2001)

# TITLE

JOURNAL  
COMMENT

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,  
Matsuda, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.  
and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
Sugahara, Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.

Location/Qualifiers

```

source
1. .357
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="G630046N07"
/sex="male"
/tissue_type="intestinal mucosa"
/dev_stage="adult"
/clone_lib="RIKEN full-length enriched, adult male
intestinal mucosa"

ORIGIN
Query Match 79.1%; Score 18.2; DB 2; Length 357;
Best Local Similarity 87.0%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
66 TGAGAAGGCAGTAGGAGCTGAG 44

RESULT 12
CG682385
LOCUS
DEFINITION
CG682385
377 bp DNA linear GSS 08-OCT-2003
ZMMBBC0143P02f ZMMBBC (EcoRI) Zea mays genomic clone ZMMBBC0143P02
5', genomic survey sequence.
ACCESSION
CG682385
VERSION
CG682385.1 GI:37579222
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE
1 (bases 1 to 377)
Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,
Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.
Sequencing of the maize genome at PGIR (2003c)
Unpublished (2003)
Contact: Bharti,A.K.
Dr.Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: T7
Class: BAC ends
High quality sequence start: 102.
FEATURES
source
Location/Qualifiers
1. .377
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="ZMMBBC0143P02"
/lab_host="E. coli DH10B"
/clone_lib="ZMMBBC (EcoRI)"
/notes="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN
Query Match 79.1%; Score 18.2; DB 10; Length 377;
Best Local Similarity 87.0%; Pred. No. 2.1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
|||||
55 TGAGATGGCAGTAGAAGGCTAAG 77

RESULT 13

```

CF982181/c  
LOCUS  
DEFINITION

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

FEATURES  
source

ORIGIN

Query Match 79.1%; Score 18.2; DB 7; Length 429;  
Best Local Similarity 87.0%; Pred. No. 2.1e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||||  
Db 225 TGAGAAGGCAGTAGGAGCAGAG 203

RESULT 14  
CK100703  
LOCUS  
DEFINITION

CF982181 429 bp mRNA linear EST 18-DEC-2003  
maj79a02.yl McCarrey Eddy 18 day preleptotene spermatocytes Mus  
musculus cDNA clone IMAGE:7032076 5' similar to TR:Q12706 Q12706  
PROBABLE SERINE/THREONINE PROTEIN KINASE PSK1 , mRNA sequence.  
CF982181  
CF982181.1 GI:38515230  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Mus.  
1 (bases 1 to 429)  
McCarrey,J., Eddy,M., Marra,M., Hillier,L., Clifton,S., Pape,D.,  
Martin,J., Wylie,T., Dante,M., Bowers,Y., Theising,B., Gibbons,M.,  
Ritter,E., Tagareishvili,R., Ronko,I., Maguire,L., Kennedy,S.,  
Bennett,J., Waterston,R. and Wilson,R.  
NIHES Mouse  
Unpublished (2002)  
Contact: McCarrey/Eddy NIHES Mouse  
NIHES Mouse  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
Library constructed and donated by J. McCarrey, Ph.D. (Southwest  
Foundation for Biomedical Research, Dept. of Genetics) - excision  
done by E.M. Eddy, Ph.D. (National Institutes of Health, National  
Institute of Environmental Health Sciences).  
Possible reversed clone: similarity on wrong strand  
Seq primer: Primer name ambiguous.

Location/Qualifiers  
1. .429  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:7032076"  
/sex="male"  
/tissue\_type="18-day preleptotene spermatocytes"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="McCarrey Eddy 18 day preleptotene  
spermatocytes"  
/note="Organ: testis; Vector: pBluescript SK+  
(Stratagene); Site 1: XhoII; Site 2: EcoRI; cDNA oligo  
dT-primed [5'-(GA)10-ACTAGTCGAGTTT-TTTT-3'] and  
directionally cloned using 5' linkers 5'-AATTGCGCAGCAG-3'  
and 5'-CTCGTGGCG-3'. Size selection of >400bp material  
gives average insert size ranging from 1-2 kb. Library was  
mass excised (from lambda-UniZAP-XR) and resulting  
single-stranded phagemids were prepped and transformed into  
DH10B. Library constructed and donated by J. McCarrey,  
Ph.D. (Southwest Foundation for Biomedical Research, Dept.  
of Genetics); excision done by E.M. Eddy, Ph.D. (National  
Institutes of Health, National Institute of Environmental  
Health Sciences)."

CK100703 433 bp mRNA linear EST 01-DEC-2003  
C067P49.5Pr Populus strain T89 leaves Populus tremula x Populus  
tremuloides cDNA clone C067P49 5', mRNA sequence.



CK100703  
 CK100703.1 GI:38585028  
 EST.  
 SOURCE  
 ORGANISM  
 Populus tremula x Populus tremuloides  
 Populus tremula x Populus tremuloides  
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Malpighiales; Salicaceae; Populus.  
 1 (bases 1 to 433)  
 Sterky, F., Bhalerao, R.R., Unneberg, P., Segerman, B., Nilsson, P.,  
 Brunner, A.M., Charbonnel-Campaa, L., Lindvall, J.U., Tandré, K.,  
 Strauss, S.H., Sundberg, B., Gustafsson, P., Uhlen, M., Bhalerao, R.P.,  
 Nilsson, O., Sandberg, G., Karlsson, J., Lundberg, J., and Jansson, S.  
 A Populus EST resource for plant functional genomics  
 Proc. Natl. Acad. Sci. U.S.A. 101 (38), 13951-13956 (2004)  
 15353603  
 Other ESTs: C067P49U  
 Contact: Bo Segerman  
 Umea Plant Science Center, Department of Plant Physiology  
 Umea University  
 901 87 Umea, Sweden  
 Tel: +46 90 786 5279  
 Fax: +46 90 786 6676  
 Email: bo.segerman@plantphys.umu.se.  
 FEATURES  
 source  
 1. .433  
 /organism="Populus tremula x Populus tremuloides"  
 /mol\_type="mRNA"  
 /strain="T89"  
 /db\_xref="taxon:47664"  
 /clone="C067P49"  
 /tissue\_type="leaf"  
 /clone\_lib="Populus strain T89 leaves"  
 ORIGIN  
 Query Match 79.1%; Score 18.2; DB 7; Length 433;  
 Best Local Similarity 87.0%; Pred. No. 2.1e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGAGAGGCGAGTAGAAGCTTAG 23  
 |||||  
 DB 294 TGAGAGGCGAGCAAAAGCGTAG 316  
 RESULT 15  
 BH025854/c  
 LOCUS  
 DEFINITION  
 RPCI-24-24607.TJ RPCI-24 Mus musculus genomic clone RPCI-24-24607,  
 genomic survey sequence.  
 ACCSSION  
 VERSION  
 BH025854.1 GI:14789318  
 KEYWORDS  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 531)  
 Zhao, S., Nierman, W., Malek, J., Shatsman, S., Akinret, B., Levins, M.,  
 Teagave, G., Geer, K., Krol, M., Shvartsbeyn, A., Gebregeorgis, E.,  
 Russell, D., de Jong, P., and Fraser, C.M.  
 Mouse BAC End Sequences from Library RPCI-24  
 Unpublished (1999)  
 Other GSSs: RPCI-24-24607.TV  
 Contact: Shaving Zhao  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-24. For BAC  
 library availability, please contact Pieter de Jong  
 (pdejong@mail.cho.org). Clones may be purchased from BACPAC  
 Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end  
 page: [http://www.tigr.org/tdb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)  
 Plate: 246 row: O column: 7  
 Seq primer: SP6  
 Class: BAC ends.  
 FEATURES  
 source  
 1. .531  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="RPCI-24-24607"  
 /sex="Male"  
 /cell\_type="Spleen/Brain"  
 /clone\_lib="RPCI-24"  
 /note="Vector: pTARBAC1; Site 1: BamHI; Site 2: BamHI;  
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The  
 library was cloned in the pTARBAC1 cloning vector at the  
 BamHI sites using MboI partially digested male C57BL/6J  
 DNA."  
 ORIGIN  
 Query Match 79.1%; Score 18.2; DB 9; Length 531;  
 Best Local Similarity 87.0%; Pred. No. 2.2e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGAGAGGCGAGTAGAAGCTTAG 23  
 |||||  
 DB 254 TGAGAGGCGAGTAAAGATTAG 232  
 RESULT 16  
 BU814643  
 LOCUS  
 DEFINITION  
 N031C09 Populus bark cdna library Populus tremula x Populus  
 tremuloides cdna 5 prime, mRNA sequence.  
 ACCSSION  
 VERSION  
 BU814643.1 GI:23972384  
 KEYWORDS  
 SOURCE  
 EST.  
 ORGANISM  
 Populus tremula x Populus tremuloides  
 Populus tremula x Populus tremuloides  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids I; Malpighiales; Salicaceae; Populus.  
 1 (bases 1 to 544)  
 Unneberg, P., Bhalerao, R.R., Jansson, S., and Sterky, F.  
 The poplar tree transcriptome: Analysis of expressed sequence tags  
 from multiple libraries  
 Unpublished (2002)  
 Contact: BHALERAO RUPALI R.  
 Umea Plant Science Center  
 Department of Plant Physiology  
 University of Umea, 901 87 Umea, Sweden  
 Tel: +46 90 786 5279  
 Fax: +46 90 786 6676  
 Email: rupali.bhalerao@plantphys.umu.se.  
 FEATURES  
 source  
 1. .544  
 /organism="Populus tremula x Populus tremuloides"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:47664"  
 /tissue\_type="bark"  
 /clone\_lib="Populus bark cdna library"  
 ORIGIN  
 Query Match 79.1%; Score 18.2; DB 5; Length 544;  
 Best Local Similarity 87.0%; Pred. No. 2.2e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGAGAGGCGAGTAGAAGCTTAG 23  
 |||||  
 DB 79 TGAGAGGCGAGCAAAAGCGTAG 101

```

RESULT 17
CA929386      567 bp  mRNA  linear  EST 30-DEC-2002
LOCUS         MTU2CA.P13.A09 Aspen apex cDNA Library Populus tremuloides cDNA,
DEFINITION    mRNA sequence.
ACCESSION     CA929386
VERSION       CA929386.1 GI:27417867
KEYWORDS      EST.
SOURCE        Populus tremuloides (quaking aspen)
ORGANISM      Populus tremuloides
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
              rosids; eurosids I; Malpighiales; Salicaceae; Saliceae; Populus.
REFERENCE     1 (bases 1 to 567)
AUTHORS       Ranjan,P., Kao,Y.-Y., Harding,S.A., Jiang,H., Joshi,C.P. and
              Tsai,C.-J.
TITLE         Expressed sequence tags from Aspen
JOURNAL       Unpublished (2003)
COMMENT       Contact: Tsai C-J
              Plant Biotech Research Center
              Michigan Technological University, School of Forest Resources &
              Environmental Science
              1400 Townsend Drive, Houghton, MI 49931-1295, USA
              Tel: 906 487 2914
              Fax: 906 487 2915
              Email: chtsai@mtu.edu.

FEATURES             Location/Qualifiers
     source           1..567
                     /organism="Populus tremuloides"
                     /mol_type="mRNA"
                     /db_xref="taxon:3693"
                     /clone_lib="Aspen apex cDNA Library"
                     /notes="Organ: apex"

ORIGIN
Query Match       79.1%; Score 18.2; DB 6; Length 567;
Best Local Similarity 87.0%; Pred. No. 2.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGGCTTAG 23
    |||||
Db 42 TGAGAAGGCAGCAAAAGCGTAG 64

RESULT 18
CF105022/c      571 bp  mRNA  linear  EST 22-JUL-2003
LOCUS           mai93c06.y1 McCarrey Eddy round spermatid Mus musculus cDNA clone
DEFINITION      IMAGE:6399754 5' similar to TR:Q12706 Q12706 PROBABLE
                SERINE/THREONINE PROTEIN KINASE PSK1 ;, mRNA sequence.
ACCESSION       CF105022
VERSION         CF105022.1 GI:33144089
KEYWORDS        EST.
SOURCE          Mus musculus (house mouse)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
                Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE       1 (bases 1 to 571)
AUTHORS         McCarrey,J., Eddy,M., Marra,M., Hillier,L., Clifton,S., Pape,D.,
                Martin,J., Wylie,T., Dante,M., Bowers,Y., Theising,B., Gibbons,M.,
                Ritter,B., Tsagarisvili,R., Ronko,I., Maguire,L., Kennedy,S.,
                Bennett,J., Waterston,R. and Wilson,R.
TITLE           Unpublished (2002)
JOURNAL         Contact: McCarrey/Eddy NIEHS Mouse
COMMENT         NIEHS Mouse
                Washington University School of Medicine
                4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
                Tel: 314 286 1800
                Fax: 314 286 1810

```

```

Email: est@watson.wustl.edu
Library constructed and donated by J. McCarrey, Ph.D. (Southwest
Foundation for Biomedical Research, Dept. of Genetics) - excision
done by E.M. Eddy, Ph.D. (National Institutes of Health, National
Institute of Environmental Health Sciences).
Possible reversed clone: similarity on wrong strand
MGI:2497197
Seq primer: Primer name ambiguous.
              Location/Qualifiers
     source           1..571
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strains="CD-1"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6399754"
                     /sex="male"
                     /tissue_type="round spermatids, pooled from multiple mice"
                     /dev_stage="60 day"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="McCarrey Eddy round spermatid"
                     /note="Organ: testis; Vector: pBluescript SK+
                     (Stratagene); Site 1: XhoII; Site 2: EcoRI; cDNA oligo
                     dt-primed [5'-(GA)10-ACTAGTCGAGTTTTTTTTT-3'] and
                     directionally cloned using 5' linkers 5'-AATTCGACAG-3'
                     and 5'-CTCGTCGCG-3'. Size selection of >400bp material
                     gives average insert size ranging from 1-2 kb. Library was
                     mass excised (from lambda-UniZAP-XR) and resulting
                     single-stranded phagemids were prepped and tranformed
                     into DH10B. Library contains 98.5% recombinants.
                     References: J. Androl. 20:635-639 and Gene 25:263-269.
                     Library constructed and donated by J. McCarrey, Ph.D.
                     (Southwest Foundation for Biomedical Research, Dept. of
                     Genetics); excision done by E.M. Eddy, Ph.D. (National
                     Institutes of Health, National Institute of Environmental
                     Health Sciences). Original lambda-based library is
                     available through ATCC, catalog #63423."

ORIGIN
Query Match       79.1%; Score 18.2; DB 6; Length 571;
Best Local Similarity 87.0%; Pred. No. 2.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGGCTTAG 23
    |||||
Db 424 TGAGAAGGCAGTAGGAGCAGAG 402

RESULT 19
BQ322034/c      601 bp  mRNA  linear  EST 17-MAY-2002
LOCUS           PM2-CT0265-120301-006-f06 CT0265 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BQ322034
ACCESSION       BQ322034
VERSION         BQ322034.1 GI:20931101
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
ORGANISM        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
                Homiidae; Homo.
REFERENCE       1 (bases 1 to 601)
AUTHORS         Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
                Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
                Goldman,G.H., Carvalho,A.F., Matsukuma,A., Balg,G.S., Simpson,D.H.,
                Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
                O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
                Simpson,A.J.
TITLE           Shotgun sequencing of the human transcriptome with ORF expressed
                sequence tags
JOURNAL         Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT         10737800
                Contact: Simpson A.J.G.
                Laboratory of Cancer Genetics
                Ludwig Institute for Cancer Research

```



SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.

REFERENCE 1 (bases 1 to 627)  
 AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.

TITLE The dog genome: survey sequencing and comparative analysis  
 JOURNAL Science 301 (5641), 1898-1903 (2003)  
 PUBMED 14512627

COMMENT Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirknes@tigr.org  
 Class: shotgun.

FEATURES Location/Qualifiers  
 source 1..627

/organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"  
 /db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from  
 peripheral Blood"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 627;  
 Best Local Similarity 87.0%; Pred. No. 2.3e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGCGCAGTAGAAGCTTAG 23  
 |||||  
 DB 132 TGAGAAGCGCAGTTGAACGCTTTG 110

RESULT 22  
 CE426147/c  
 LOCUS tigr-gss-dog-17000362829894 Dog Library Canis familiaris genomic,  
 DEFINITION genomic survey sequence.

ACCESSION CE426147  
 VERSION CE426147.1 GI:36695469  
 KEYWORDS GSS.

SOURCE Canis familiaris (dog)  
 ORGANISM Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.

REFERENCE 1 (bases 1 to 640)  
 AUTHORS Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.

TITLE The dog genome: survey sequencing and comparative analysis  
 JOURNAL Science 301 (5641), 1898-1903 (2003)  
 PUBMED 14512627

COMMENT Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-838-0200  
 Fax: 301-838-0208  
 Email: ekirknes@tigr.org  
 Class: shotgun.

FEATURES Location/Qualifiers  
 source 1..640  
 /organism="Canis familiaris"  
 /mol\_type="genomic DNA"  
 /strain="Standard Poodle"

/db\_xref="taxon:9615"  
 /clone\_lib="Dog Library"  
 /note="Site 1: BstXI; Libraries were prepared from  
 peripheral Blood"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 640;  
 Best Local Similarity 87.0%; Pred. No. 2.3e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGCGCAGTAGAAGCTTAG 23  
 |||||  
 DB 444 TGAGAAGCGCAGTATAAACTTGG 422

RESULT 23  
 BZ984463/c

LOCUS PUF1180TD ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBTa350M15,  
 DEFINITION genomic survey sequence.

ACCESSION BZ984463  
 VERSION BZ984463.1 GI:29219281  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 644)  
 AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
 Bennetzen,J.

TITLE Maize Genomics Consortium  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: PUF1180TB  
 Contact: Cathy Whitelaw

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: sheared ends.

FEATURES Location/Qualifiers  
 source 1..644  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBTa350M15"  
 /clone\_lib="ZM\_0.6\_1.0\_KB"  
 /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
 CoT selected genomic DNA library"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 9; Length 644;  
 Best Local Similarity 87.0%; Pred. No. 2.3e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGCGCAGTAGAAGCTTAG 23  
 |||||  
 DB 39 TGAGAAGCGCAGTAGAAGCTTGC 17

RESULT 24  
 CX179363/c

LOCUS CX179363  
 DEFINITION euramericana Populus euramericana cdNA, mRNA sequence.  
 EST.

ACCESSION CX179363  
 VERSION CX179363.1 GI:56826787  
 KEYWORDS EST.  
 SOURCE Populus euramericana

CX179363 671 bp mRNA linear EST 28-DEC-2004  
 D07\_45-16\_07.ab1 leaf inoculated with Marssonina pathogen of Populus  
 euramericana

```

ORGANISM
Populus euramericana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids I; Malpighiales; Salicaceae; Salicaceae; Populus.
REFERENCE
AUTHORS
1 (bases 1 to 671)
Huang,M., Zhang,X., Zhuge,Q., Zhou,Y., Zhou,H., Zhang,L., Guan,Y.,
Zhang,B. and Wang,M.
TITLE
Identification and Expression Analysis of EST-based Genes in the
Leaf of poplar
JOURNAL
Unpublished (2004)
COMMENT
Contact: Minren Huang
Key Laboratory Tree Genetic Engineering of Nanjing Forestry
University
Nanjing Forestry University
Longpan road 9#, Nanjing, Jiangsu, P.R.China
Tel: 086-025-85427412
Fax: 086-025-85427412
Email: mrhuang@njfu.edu.cn
POLYA=Yes.

FEATURES
source
1..671
/organism="Populus euramericana"
/mol_type="mRNA"
/db_xref="taxon:106131"
/clone_lib="leaf inoculated with Marssonina pathogen of
Populus euramericana"
/notes="vector: pBK-CMV, Site_1: EcoRI; Site_2: XhoI"

ORIGIN
Query Match 79.1%; Score 18.2; DB 8; Length 671;
Best Local Similarity 87.0%; Pred. No. 2.3e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
|||||
Db 425 TGAGAGGCAGCAAAAGCGTAG 403

RESULT 25
SP6106024/c
LOCUS
DEFINITION
Equus caballus GSS, BAC clone CH241-106024, SP6 end sequence,
genomic survey sequence.
ACCESSION
CR008091
VERSION
CR008091.1 GI:68130624
KEYWORDS
GSS; genomic survey sequence.
SOURCE
Equus caballus (horse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Perissodactyla; Equidae; Equus.
REFERENCE
AUTHORS
1 (bases 1 to 731)
Leeb,T., Binns,M.M., de Jong,P.J., Berg,C., Conrad,A., Jarek,M.,
Loehner,T.H., Nordliek,G., Severitt,S., Scharfe,S.,
Schindewolf,C., Schrader,F., Thies,S. and Bloecker,H.
TITLE
Analysis of horse BAC sequences
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 731)
Jarek,M.
TITLE
Direct Submission
AUTHORS
Submitted (10-JUN-2005) of Genome Analysis, German Research Centre
for Biotechnology, Mascheroder Weg 1, Braunschweig D-38124, Germany
JOURNAL
Location/Qualifiers
1..731
/organism="Equus caballus"
/mol_type="genomic DNA"
/db_xref="taxon:9796"
/clone="CH241-106024"
/clone_lib="CHORI-241"
/notes="sequenced with SP6 primer"

FEATURES
source
1..731
/organism="Equus caballus"
/mol_type="genomic DNA"
/db_xref="taxon:9796"
/clone="CH241-106024"
/clone_lib="CHORI-241"
/notes="sequenced with SP6 primer"

ORIGIN
Query Match 79.1%; Score 18.2; DB 11; Length 731;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
|||||
Db 471 TGAGAGGCAGTAGAAGCAGAG 449

RESULT 27
BI737502
LOCUS
DEFINITION
603358184F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5365317 5',
mRNA sequence.
ACCESSION
BI737502
VERSION
BI737502.1 GI:15714515
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

```

```

Matches
20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
|||||
Db 360 TGAGATGTAGTAGAAGCTGAG 338

RESULT 26
BU946020/c
LOCUS
DEFINITION
AGENCOURT10563999 NIH_MGC_169 Mus musculus cDNA clone
IMAGE:6740737 5', mRNA sequence.
ACCESSION
BU946020
VERSION
BU946020.1 GI:24134839
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 732)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. Jonathan Kuo, NIMH
CDNA Library Preparation: Michael Brownstein Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC3076 row: a column: 24
High quality sequence stop: 571.
Location/Qualifiers
1..732
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:6740737"
/lab_host="DHI0B (T1-phage-resistant)"
/clone_lib="NIH_MGC_169"
/notes="Organ: Testicles; Vector: pDNR-LIB; Site_1: SfiI
(gccattatggcc); Site_2: SfiI (ggcgctcgcc); CDNA made
by oligo-dT priming and directionally cloned. 5' and 3'
adaptors were used in cloning as follows:
5'-AAGCAGTGGTATCAACGAGTCGCGCATTCAGCCGG-3' and
5'-ATTCTAGAGCCGAGCGCGCATATG-dT(30)NN-3'. Full-length
enriched library was constructed using the Clontech
Creator SMART kit and size-selected to contain the 0.5 kb
size fraction. Library created in the laboratory of M.
Brownstein (NIMH, NIH). Note: this is a NIH_MGC Library."

ORIGIN
Query Match 79.1%; Score 18.2; DB 5; Length 732;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCAGTAGAAGCTTAG 23
|||||
Db 471 TGAGAGGCAGTAGAAGCAGAG 449

RESULT 27
BI737502
LOCUS
DEFINITION
603358184F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5365317 5',
mRNA sequence.
ACCESSION
BI737502
VERSION
BI737502.1 GI:15714515
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

```

```
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 740)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: The Cepko Laboratory
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM11929 row: h column: 22
              High quality sequence stop: 738.
              Location/Qualifiers
                1..740
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:5365317"
                  /tissue_type="retina"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NIH MGC 94"
                  /note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
                  Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                  Average insert size 3.3 kb. Library enriched for
                  full-length clones and constructed by Life Technologies.
                  Note: this is a NIH_MGC Library."

ORIGIN
Query Match      79.1%; Score 18.2; DB 3; Length 740;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23
    |||||
DB 696 TGAGAAGGCAGTAGAAGCAGAG 718

RESULT 28
CO431201
LOCUS          772 bp mRNA linear EST 06-JUL-2004
DEFINITION    UI-W-HX0-csd-e-18-0-UI.r1 NIH BMAP_HX0 Mus musculus cDNA clone
ACCESSION     CO431201
VERSION       CO431201.1 GI:49677495
KEYWORDS      Mus musculus (house mouse)
SOURCE        Mus musculus
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
              Sciurognathi; Muroidea; Muridae; Murinae; Mus.
              1 (bases 1 to 772)
REFERENCE     NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE        Unpublished (1999)
JOURNAL       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Dr. James Lin University of Iowa
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Distribution information can be found at
              http://genome.uiowa.edu/distribution/mouseefl.html
              This clone was contributed by the Brain Molecular Anatomy Project
              (BMAP)
              Seq primer: pYX-5.
              Location/Qualifiers

FEATURES
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 772)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL       Unpublished (1999)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: The Cepko Laboratory
              cDNA Library Preparation: Life Technologies, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: LLAM11929 row: h column: 22
              High quality sequence stop: 738.
              Location/Qualifiers
                1..772
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:30685457"
                  /tissue_type="whole eye"
                  /dev_stage="newborn (1, 5, 15 days) and embryonic (15, 16,
                  17, 18 dpc)"
                  /lab_host="DH10B (T1 phage resistant)"
                  /clone_lib="NIH BMAP_HX0"
                  /note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
                  Site 2: Not I; The library was constructed according
                  to Bontade, Lennon and Soares, Genome Research, 6:791-806,
                  1996. Denatured RNA was size fractionated on a 1% agarose
                  gel. First strand cDNA synthesis was primed with oligo-dT
                  primer containing a Not I site. Double strand cDNA was
                  size selected according to mRNA size fraction, ligated
                  with EcoR I adaptor, digested with NotI and then cloned
                  directionally into pYX-Asc vector. The library tag
                  sequence located between the Not I site and the polyA tail
                  is AATAAATACG. This library was created for the University
                  of Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the
                  Developing Mouse Nervous System', supported by National
                  Institute of Mental Health (NIMH)."
```

```
source
1..772
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30685457"
/tissue_type="whole eye"
/dev_stage="newborn (1, 5, 15 days) and embryonic (15, 16,
17, 18 dpc)"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_HX0"
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according
to Bontade, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AATAAATACG. This library was created for the University
of Iowa Brain Anatomy Project (BMAP); 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH)."
```

ORIGIN

Query Match 79.1%; Score 18.2; DB 7; Length 772;  
Best Local Similarity 87.0%; Pred. No. 2.4e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAAGGCAGTAGAAGCTTAG 23  
|||  
DB 610 TGAGAAGGCAGTAGAAGCAGAG 632

RESULT 29  
BU466365  
LOCUS 775 bp mRNA linear EST 30-NOV-2002  
DEFINITION 603268708F1 CSEQRBN20 Gallus gallus cDNA clone CHEST241o2 5', mRNA  
sequence.  
ACCESSION BU466365  
VERSION BU466365.1 GI:25959942  
KEYWORDS Gallus gallus (chicken)  
SOURCE Gallus gallus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 775)  
REFERENCE Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curt. Biol. 12 (22), 1965-1969 (2002)  
12445392  
COMMENT Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology  
(UMIST)  
PO Box 88, Manchester, M60 1QD, UK  
Tel: 01612008930  
Fax: 01612360409  
Email: Simon.Hubbard@umist.ac.uk.  
Location/Qualifiers  
1..775  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Layer and broiler"  
/db\_xref="taxon:9031"  
/clone="CHEST241o2"  
/sex="Male and female"  
/tissue\_type="Chondrocytes isolated from growth plate  
cartilage"

FEATURES  
source

/dev\_stage="adult"  
 /lab\_host="DH108"  
 /clone\_lib="CSQRN20"  
 /note="Vector: pBluescript II KS(+); Site\_1: EcoRI;  
 Site\_2: NotI; This normalized library was constructed from  
 1 million independent clones. cDNA synthesis was initiated  
 using an oligo(dT) primer, using methylated C in the first  
 strand synthesis reaction. Following this first strand  
 reaction, double-stranded cDNA was blunted, ligated to  
 NotI adapters, digested with EcoRI, size-selected, and  
 cloned into the NotI and EcoRI compatible sites of a  
 custom modified MCS of the pBluescript (KS+) vector. The  
 library was normalized in 2 rounds using conditions  
 adapted from Soares et al., PNAS (1994) 91: 9228-9232 and  
 Bonaldo et al., Genome Research 6 (1996): 791, except that  
 a significantly longer reannealing hybridization was  
 used."

## ORIGIN

Query Match 79.1%; Score 18.2; DB 5; Length 775;  
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCGCAGTAGAAGCTTAG 23  
 |||||  
 Db 623 TGAGAGGCGCAGTAGAAGCTTAG 645

RESULT 30  
CA494633/c

LOCUS CA494633 819 bp mRNA linear EST 14-NOV-2002  
 DEFINITION AGENCOURT\_10815914 NIH\_MGC\_169 Mus musculus cDNA clone  
 IMAGE:6774068 5', mRNA sequence.

ACCESSION CA494633  
 VERSION CA494633.1 GI:24957679  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidae; Muridae; Murinae; Mus.

1 (bases 1 to 819)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Jonathan Kuo, NIMH

cDNA Library Preparation: Michael Brownstein Laboratory

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LUCM3097 row: n column: 19

High quality sequence stop: 507.

Location/Qualifiers

FEATURES  
source

1..819  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6774068"  
 /lab\_host="DH10B (T1-phage-resistant)"  
 /clone\_lib="NIH MGC 169"  
 /note="Torgan: Testicles; Vector: pDNR-LIB; Site 1: SfiI  
 (ggccattatggcc); Site 2: SfiI (ggcgctcgcc); cDNA made  
 by oligo-dT priming and directionally cloned. 5' and 3'  
 adaptors were used in cloning as follows:  
 5'-AAGCAGTGTGATACGACGATGCGCATACGCCCGG-3' and  
 5'-ATTCTAGAGCGGCGGCGGCACATG-dT(30)NN-3'. Full-length  
 enriched library was constructed using the Clontech  
 Creator SMART kit and size-selected to contain the 0.5 kb  
 size fraction. Library created in the laboratory of M.

Brownstein (NIMH, NIH). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 79.1%; Score 18.2; DB 6; Length 819;  
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCGCAGTAGAAGCTTAG 23  
 |||||  
 Db 613 TGAGAGGCGCAGTAGAAGCTTAG 591

## RESULT 31

CZ717363

LOCUS CZ717363

DEFINITION

5', genomic survey sequence.

ACCESSION CZ717363

VERSION CZ717363.1 GI:71125837

KEYWORDS GSS.

SOURCE Oryza coarctata (Porteresia coarctata)

ORGANISM Oryza coarctata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 882)

REFERENCE Kim H., Collura, K., Wisotski, M., Byrne, M., Stum, D., Smart, D.,

Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C., Soderlund, C. and

Wing, R.

OMAP (Oryza Map Alignment Project) - Arizona Genomics Institute

Unpublished (2005)

Contact: Rod A. Wing

Arizona Genomics Institute

University of Arizona

Forbes Building Room 303, Tucson, AZ 85721-0036, USA

Tel: 520 626 9595

Fax: 520 621 1259

Email: rwing@genome.arizona.edu

PCR Primers

FORWARD: TAA TAC GAC TCA CTA TAG GG

BACKWARD: CAC TCA TTA GGC ACC CCA

Plate: 0041 row: I column: 15

Seq primer: TAA TAC GAC TCA CTA TAG GG

Class: BAC ends.

Location/Qualifiers

1..882

/organism="Oryza coarctata"

/mol\_type="genomic DNA"

/db\_xref="taxon:77588"

/clone="OC\_Ba0041115"

/tissue\_type="leaves"

/dev\_stage="mature"

/lab\_host="DH10B"

/clone\_lib="OC\_Ba"

/notes="Vector: pAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 79.1%; Score 18.2; DB 10; Length 882;  
 Best Local Similarity 87.0%; Pred. No. 2.4e+03;  
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAGAGGCGCAGTAGAAGCTTAG 23  
 |||||  
 Db 742 TGAGAGGCGCAGTAGAAGCTTAG 764

## RESULT 32

DT058135

LOCUS DT058135

DEFINITION

AGENCOURT\_55952926 NICHG\_XGC\_Fab xenopus laevis cDNA clone

IMAGE:8069608 3', mRNA sequence.

ACCESSION DT058135

VERSION DT058135.1 GI:72357384

DT058135 886 bp mRNA linear EST 11-AUG-2005  
 AGENCOURT\_55952926 NICHG\_XGC\_Fab xenopus laevis cDNA clone  
 IMAGE:8069608 3', mRNA sequence.  
 DT058135  
 DT058135.1 GI:72357384



```

KEYWORDS
SOURCE  Xenopus laevis (African clawed frog)
ORGANISM

REFERENCE
1 (bases 1 to 886)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Daniela S. Gerhard, Ph.D.
         Office of Cancer Genomics
         National Cancer Institute / NIH
         Bldg. 31 Rm10A07 Bethesda, MD 20892
         Email: cgapbs-r@mail.nih.gov
         Tissue Procurement: Tom Sargent
         cDNA Library Preparation: Express Genomics
         DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
         DNA Sequencing by: Agencourt Bioscience Corporation
         Clone distribution: MGC clone distribution information can be
         found through the I.M.A.G.E. Consortium/LLNL at:
         http://image.llnl.gov
         Plate: LLAM17342 row: k column: 14
         High quality sequence start: 398
         High quality sequence stop: 755.

FEATURES
source
1..886
    /organism="Xenopus laevis"
    /mol_type="mRNA"
    /db_xref="taxon:8355"
    /clone="IMAGE:8069608"
    /lab_host="DH10B Tona"
    /clone_lib="NICHG_XGC_Fab"
    /notes="Organ: fat body; Vector: pExpress-1; Site_1: EcoRV;
    Site_2: NotI; cDNA was primed using oligo-dT primer:
    5'-pGACTAGTTCTAGATCCGAGCGCGCCCC(T)25-3' and cloned into
    the EcoRV/NotI sites of pExpress-1. Size-selection 1.2kb
    resulted in an average insert size of 1.8kb. This is a
    primary library (normalized library is NICHG_XGC_FabN) and
    was constructed by Express Genomics (Frederick, MD). Note:
    this is a (http://xgc.nci.nih.gov/) Xenopus Gene
    Collection library."

ORIGIN
Query Match 79.1%; Score 18.2; DB 8; Length 886;
Best Local Similarity 87.0%; Pred. No. 2.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 742 TGAGAGGCGAGTAGAAGCTTAG 764

RESULT 33
LOCUS BC037104
DEFINITION Mus musculus cDNA sequence BC030499, mRNA (cDNA clone
IMAGE:5365317), with apparent retained intron.
BC037104
BC037104.1 GI:23331120
HTC.
Mus musculus (house mouse)
SOURCE
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1381)
AUTHORS  Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Bretfield,Y.S., Krzywinski,M.I., Skalska,U., Small,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
JOURNAL 2 (bases 1 to 1381)
PUBMED Director MGC Project.
REFERENCE Direct Submission
AUTHORS Submitted (23-AUG-2002) National Institutes of Health, Mammalian
TITLE Gene Collection (MGC), Cancer Genomics Office, National Cancer
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nigr.nih.gov
Akhrez,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granitz,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaepi,R.,
Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stantripo,S., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Place: 81 Row: 0 Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: Genomescan gene prediction
This clone has the following problem: retained intron.

FEATURES
source
1..1381
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /clone="IMAGE:5365317"
    /tissue_type="Eye, retina, mouse strain C57Bl/6"
    /clone_lib="NIH_MGC_94"
    /lab_host="DH10B"
    /note="Vector: pCMV-SPORT6"

ORIGIN
Query Match 79.1%; Score 18.2; DB 4; Length 1381;
Best Local Similarity 87.0%; Pred. No. 2.6e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTTAG 23
    ||||| ||||| ||||| ||||| |||||
Db 711 TGAGAGGCGAGTAGAAGCGAG 733

RESULT 34
LOCUS BC048862
1647 bp mRNA linear HTC 03-AUG-2004

```



```
DEFINITION Mus musculus cDNA sequence BC030499, mRNA (cDNA clone
IMAGE:5367114), with apparent retained intron.
ACCESSION BC048862
VERSION BC048862.1 GI:29179567
KEYWORDS Mus musculus (house mouse)
SOURCE HTc.
ORGANISM Mus musculus

REFERENCE
AUTHORS Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Donald,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
Worley,K.C., Hale,S., Garcia,A.M., Madan,A., Young,A.C., Shevchenko,Y.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
JOURNAL PUBLISHED 2 (bases 1 to 1647)
REFERENCE DIRECTOR MGC Project.
AUTHORS Direct Submission
TITLE Submitted (14-WAR-2003) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapsb-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Email: cgapsb-remail.nih.gov
Akhter,N., Ayale,K., Becketrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: TRAK Plate: 99 Row: k Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomesScan gene prediction
This clone has the following problem: retained intron.
Location/Qualifiers
1..1647
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5367114"
/tissue_type="Eye, retina, mouse strain C57Bl/6"
FEATURES
source
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
ORIGIN
Query Match 79.1%; Score 18.2; DB 4; Length 1647;
Best Local Similarity 87.0%; Pred. No. 2.7e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TCAGGAAGGCAGTAGAAGAGCTTAG 23
Db 878 TCAGGAAGGCAGTAGGAGCAGAG 900
RESULT 35
BC038223
LOCUS BC038223
DEFINITION Homo sapiens DEAH (Asp-Clu-Ala-His) box polypeptide 8, mRNA (cDNA
clone IMAGE:4215266), with apparent retained intron.
ACCESSION BC038223
VERSION BC038223.1 GI:23468302
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 6244)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shennen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Donald,M.F., Casavant,T.L.,
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
JOURNAL PUBLISHED 2 (bases 1 to 6244)
REFERENCE DIRECTOR MGC Project.
AUTHORS Direct Submission
TITLE Submitted (30-SEP-2002) National Institutes of Health, Mammalian
JOURNAL Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapsb-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayale,K., Becketrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
Tsurgoun,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov
Series: TRAK Plate: 99 Row: k Column: 13
This clone was selected for full length sequencing because it
passed the following selection criteria: GenomesScan gene prediction
This clone has the following problem: retained intron.
Location/Qualifiers
1..1647
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:5367114"
/tissue_type="Eye, retina, mouse strain C57Bl/6"
```

McDowell, J., Pearson, R., Stantripop, S., Thomas, P. J., Touchman, J. W., Tsugeon, C., Vogt, J. L., Walker, M. A., Wetherby, K. D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E. D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 64 Row: a Column: 12  
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis  
This clone has the following problem: retained intron.

#### FEATURES

Location/Qualifiers  
1..6244  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4215266"  
/tissue type="Brain, anaplastic oligodendroglioma with lp/19q loss"  
/clone lib="NCI CGAP\_Brn67"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"

#### ORIGIN

Query Match 79.1%; Score 18.2; DB 4; Length 6244;  
Best Local Similarity 87.0%; Pred. No. 3.4e+03;  
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAGAGGCGAGTAGAAGCTTAG 23

Db 3421 TGAGAGGAGAAATAGAAAGCTTAG 3443

RESULT 36  
BH868412/C

LOCUS

DEFINITION  
hk25c06.x7 WGS-Zmay5F (JM107 adapted methyl filtered) Zea mays  
Genomic clone hk25c06 5', genomic survey sequence.

ACCESSION  
BH868412

VERSION  
GSS.

SOURCE  
Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE  
1 (bases 1 to 272)

AUTHORS

Rabinowicz, P.D., O'Shaughnessy, A.L., Balija, V., Dedhia, N., Katzenburger, F., King, L., Miller, B., Muller, S., Nascimento, L., Zutavern, T., McCombie, W.R. and Martienssen, R.A.,  
Genomic shotgun sequences from Zea mays (methyl-filtered)  
Unpublished (2002)

TITLE

JOURNAL

COMMENT

Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: [mccombie@cshl.org](mailto:mccombie@cshl.org)

Plate: hk25 row: c column: 06

Seq primer: -21M13UnivRev

Class: shotgun

High quality sequence stop: 272.

FEATURES

source

Location/Qualifiers

1..272

/organism="Zea mays"

/mol\_type="genomic DNA"

/cultivar="B73"

/db\_xref="taxon:4577"

/clone="hk25c06"

/lab\_host="JM107 or DH5a"

/clone lib="WGS-Zmay5F (JM107 adapted methyl filtered)"

/note="Organ: Immature ears; Site\_1: Xba I; Site\_2: Xba I;

#### ORIGIN

Query Match 77.4%; Score 17.8; DB 9; Length 272;  
Best Local Similarity 90.5%; Pred. No. 3e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCGAGTAGAAGCTTA 22

Db 133 GAGAGGCGAGTAGAAGCTTA 113

RESULT 37

LOCUS

DEFINITION  
CO747995 314 bp mRNA linear EST 02-AUG-2004  
Library Sarcocystis neuropa cDNA 5', mRNA sequence.

ACCESSION  
CO747995

KEYWORDS  
EST.

SOURCE

ORGANISM

Sarcocystis neuropa  
Sarcocystis neuropa  
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
Sarcocystidae; Sarcocystis.

REFERENCE  
1 (bases 1 to 314)

AUTHORS

Howe, D.K., Stamper, S., Tang, K., Sibley, L.D., Clifton, S., Marra, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., McCann, R., Blistain, A., Bennett, J., Schmitt, A., Ronko, I., Tsagareishvili, R., Fedele, M., Belaygorod, L., Franklin, C., Carr, L.M., Grow, A., Maguire, L., Wadkins, J., Richey, J., Waterston, R. and Wilson, R.  
Sarcocystis neuropa EST project  
Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Daniel K. Howe  
Sarcocystis neuropa EST project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

Contact Daniel K. Howe ([dkhowe2@pop.uky.edu](mailto:dkhowe2@pop.uky.edu)) for further information relating to organism, libraries, or clone availability.  
Sequenced by Washington University Genome Sequencing Center  
Seq primer: -21UPot  
High quality sequence stop: 314.

FEATURES

source

Location/Qualifiers

1..314

/organism="Sarcocystis neuropa"

/mol\_type="mRNA"

/strain="SN4"

/db\_xref="taxon:42890"

/dev\_stage="merozoite"

/lab\_host="GC10"

/clone lib="Sarcocystis neuropa merozoite UK CSN4 1.cDNA library"

/note="Vector: pDNR-LIB; Site\_1: GGCGCGCTCGGC; Site\_2: GGCCATTACGCC; Library constructed by: Daniel K. Howe and Michelle R. Yeagan Total RNA was isolated from culture-derived merozoites of Sarcocystis neuropa strain SN4. cDNA was synthesized using the template-switching and long-distance PCR method (SMART cDNA library construction kit, BD Biosciences). The amplified cDNA fragments were digested with SfiI, size fractionated, and ligated into SfiI-digested pDNR-LIB vector."

#### ORIGIN

Query Match

77.4%; Score 17.8; DB 7; Length 314;

Best Local Similarity 90.5%; Pred. No. 3.1e+03; Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GAGAGGCAGTAGAAGCTTA 22  
|||||  
Db 40 GAGAGGCAGTAGAGGCTTA 60

RESULT 38  
CV629435 367 bp mRNA linear EST 25-OCT-2004  
LOCUS Mdfrt310121.y1 Mdfrt Malus x domestica cDNA clone Mdfrt310121 5',  
DEFINITION mRNA sequence.  
ACCESSION CV629435  
VERSION CV629435.1 GI:54622299  
KEYWORDS EST.  
SOURCE Malus x domestica  
ORGANISM Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.  
1 (bases 1 to 367)  
Korban, S., Vodkin, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,  
Aldwinckle, H., Malnov, M., Carroll, N., Goldsbrough, P., Orvis, K.,  
Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,  
Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Ronko, I.,  
Tsagarashvili, R., Kennedy, S., Waterston, R., and Wilson, R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)  
Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library materials provided by: Schuyler S. Korban Library  
Constructed by: K. Gasic Library sequenced by: Washington  
University Genome Sequencing Center  
WashU EST name: aal7leil.y1  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1..367  
/organism="Malus x domestica"  
/mol\_type="mRNA"  
/cultivar="GoldRush"  
/db\_xref="taxon:3750"  
/clone="Mdfrt310121"  
/tissue\_type="Fruit"  
/lab\_host="DH10B ampicillin resistant"  
/clone\_lib="Mdfrt"  
/notes="vector: pBluescript II SK (+); Site 1: NotI;  
Site 2: EcoRI; Total RNA was extracted separately from  
each stage (young fruitlet (<1cm), young fruitlet (1 cm  
dia.), young fruitlet (12cm dia.), maturing fruit I,  
maturing fruit II, mature fruit), using the 'pine tree'  
method. Poly(A)+mRNA was isolated twice from total RNA  
from each stage using the Oligotex Direct mRNA kit  
(Qiagen). mRNA was reverse transcribed into double  
stranded cDNA using a modified oligo(dT) primer with an  
identifying tag sequence (see table below). cDNA's from  
different stages were pooled in equal amounts before  
adaptor ligation. Tag identification when sequencing from  
5', end: Stage 1 (young fruitlet) insert 18(A)TCGTG; Stage  
2 (young fruitlet 1cm dia) insert 18(A)TCGTG; Stage 3  
(young fruitlet 12cm dia) insert 18(A)TCGTG; Stage 4  
(maturing fruit I) insert 18(A)TCGA; Stage 5 (maturing  
fruit II) insert 18(A)TCGA; Stage 6 (mature fruit) insert  
18(A)TCGG; Tag identification when sequencing from 3',  
end: Stage 1 (young fruitlet) CACGA18(T) insert; Stage 2  
(young fruitlet 1cm dia) CACGA18(T) insert; Stage 3 (young  
fruitlet 12cm dia) ACCGA18(T) insert; Stage 4 (maturing  
fruit I) TCGCA18(T) insert; Stage 5 (maturing fruit II)

TCGA18(T) insert; Stage 6 (mature fruit) ACCGA18(T)  
insert. Double stranded cDNAs were size selected (more  
than 450 bp), adapted with EcoRI adaptors at both ends  
and then digested with NotI. The cDNAs were then  
directionally cloned into EcoRI-NotI digested pBS II SK(+) phagemid vector (Stratagene). Identification of adaptors and tags in 5'-end sequenced clones:  
(Vector) . . . TAAAGCTT(End Vector)(Start  
EcoRI adaptor)GATATCGAATTCATGTTGTTGGG (End  
EcoRI adaptor)(Start Insert) . . . AAAAAAAAAAAAAAAAAA (End  
Insert) (Start Tag)TCGA(End Tag) (Start  
NotI site/Vector)GGGCGCCACCGCGG. . . The total number of  
white colony forming units (cfu) in the primary library  
before amplification was 2.1x10<sup>6</sup> cfu (colony forming  
units). The background of empty clones was less than 1%.  
Inserts ranged from 0.5kb to 4 kb, as determined by PCR.  
Purified plasmid DNA from the primary library was  
converted to single-stranded circles and used as a  
template for PCR amplification using the T7 and T3 priming  
sites flanking the cloned cDNA inserts. The purified PCR  
products, representing the entire cloned cDNA population,  
were used as a driver for normalization. Hybridization  
between the single-stranded library and the PCR products  
was carried out for 44 hours at 30C. Unhybridized  
single-stranded DNA circles were separated from hybridized  
DNA rendered partially double-stranded and electroporated  
into DH10B cells to generate the normalized library. The  
total number of clones with insert was 5.6x10<sup>6</sup> cfu.  
Background of empty clones was less than 1%."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 7; Length 367;  
Best Local Similarity 90.5%; Pred. No. 3.2e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGGCAGTAGAAGCTTAG 23  
|||||  
Db 130 AGAAGGCTGAGAAATCTTAG 150

## RESULT 39

CV630812 367 bp mRNA linear EST 25-OCT-2004  
LOCUS Mdfrt3105c20.y1 Mdfrt Malus x domestica cDNA clone Mdfrt3105c20 5',  
DEFINITION mRNA sequence.  
ACCESSION CV630812  
VERSION CV630812.1 GI:54623676  
KEYWORDS EST.  
SOURCE Malus x domestica  
ORGANISM Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.  
1 (bases 1 to 367)  
Korban, S., Vodkin, L., Liu, L., Gasic, K., Gonzales, O., Hernandez, A.,  
Aldwinckle, H., Malnov, M., Carroll, N., Goldsbrough, P., Orvis, K.,  
Clifton, S., Pape, D., Marra, M., Hillier, L., Martin, J., Wylie, T.,  
Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Ronko, I.,  
Tsagarashvili, R., Kennedy, S., Waterston, R., and Wilson, R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)  
Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library materials provided by: Schuyler S. Korban Library  
Constructed by: K. Gasic Library sequenced by: Washington  
University Genome Sequencing Center  
WashU EST name: aal88b10.y1  
Seq primer: -40UP from Gibco.

FEATURES  
source

Location/Qualifiers  
1. .367  
/organism="Malus x domestica"  
/mol\_type="mRNA"  
/cultivar="GoldRush"  
/db\_xref="taxon:3750"  
/clone="Mdftr3105c20"  
/tissue\_type="Fruit"  
/lab\_host="DH10B ampicillin resistant"  
/clone\_lib="Mdftr"  
/note="Vector: pBluescript II SK (+); Site 1: NotI;  
Site 2: EcoRII; Total RNA was extracted separately from  
each stage (young fruitlet (<1cm), young fruitlet (1 cm  
dia.), young fruitlet (12cm dia.), maturing fruit I,  
maturing fruit II, mature fruit), using the 'pine tree'  
method. Poly(A)+mRNA was isolated twice from total RNA  
from each stage using the Oligotex Direct mRNA kit  
(Qiagen). mRNA was reverse transcribed into double  
stranded cDNA using a modified oligo18(dT) primer with an  
identifying tag sequence (see table below). cDNA's from  
different stages were pooled in equal amounts before  
adaptor ligation. Tag identification when sequencing from  
5' end: Stage 1 (young fruitlet) insert 18(A)TCGTG; Stage  
2 (young fruitlet 1cm dia) insert 18(A)TCGTG; Stage 3  
(young fruitlet 12cm dia) insert 18(A)TCGTG; Stage 4  
(maturing fruit I) insert 18(A)TCGA; Stage 5 (maturing  
fruit II) insert 18(A)TCGA; Stage 6 (mature fruit) insert  
18(A)TCGTG; Tag identification when sequencing from 3',  
end: Stage 1 (young fruitlet) CAGCA18(T) insert; Stage 2  
(young fruitlet 1cm dia) CAGCA18(T) insert; Stage 3 (young  
fruitlet 12cm dia) ACCGA18(T) insert; Stage 4 (maturing  
fruit I) TCGCA18(T) insert; Stage 5 (maturing fruit II)  
TCGCA18(T) insert; Stage 6 (mature fruit) ACCGA18(T)  
insert. Double stranded cDNAs were size selected (more  
than 450 bp), adaptored with EcoRI adaptors at both ends  
and then digested with NotI. The cDNAs were then  
directionally cloned into EcoRI-NotI digested pBS II SK(+)  
phagemid vector(Stratagene). Identification of adaptors  
and tags in 5'-end sequenced clones:  
(Vector) . . .TAAGCTT(End Vector)(Start  
EcoRI adaptor)GATATCGAATTCATTGTTGGG (End  
EcoRI\_adaptor)(Start Insert). AAAAAAAAAAAAAAAAAA(End  
Insert) (Start Tag)TCGA(End Tag) (Start  
NotI site/Vector)GGCGCCACCGCG. . . The total number of  
white colony forming units (cfu) in the primary library  
before amplification was 2.1x10<sup>6</sup> cfu (colony forming  
units). The background of empty clones was less than 1%.  
Inserts ranged from 0.5kb to 4 kb, as determined by PCR.  
Purified plasmid DNA from the primary library was  
converted to single-stranded circles and used as a  
template for PCR amplification using the T7 and T3 priming  
sites flanking the cloned cDNA inserts. The purified PCR  
products, representing the entire cloned cDNA population,  
were used as a driver for normalization. Hybridization  
between the single-stranded library and the PCR products  
was carried out for 44 hours at 30C. Unhybridized  
single-stranded DNA circles were separated from hybridized  
DNA rendered partially double-stranded and electroporated  
into DH10B cells to generate the normalized library. The  
total number of clones with insert was 5.6x10<sup>6</sup> cfu.  
Background of empty clones was less than 1%."

## ORIGIN

Query Match 77.4%; Score 17.8; DB 7; Length 367;  
Best Local Similarity 90.5%; Pred. No. 3.2e+03;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AGAAGCAGGTAGAAAGCTTAG 23

|||||

Db 130 AGAAGGCTGTAGAAATCTTAG 150

|||||

RESULT 40

CV657147  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CV657147 367 bp mRNA linear EST 27-OCT-2004  
Mdftr3109hl7.y1 Mdftr Malus x domestica cDNA clone Mdftr3109hl7 5',  
mRNA sequence.  
CV657147  
CV657147.1 GI:54684533  
EST.  
Malus x domestica  
Malus x domestica  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Rosales; Rosaceae; Maloideae; Malus.  
1 (bases 1 to 367)  
Korban,S., Vodka,L., Liu,L., Gasic,K., Gonzales,O., Hernandez,A.,  
Aldwinckle,H., Malnoy,M., Carroll,N., Goldsbrough,P., Orvis,K.,  
Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T.,  
Dante,M., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Ronko,I.,  
Tsagarisvili,R., Kennedy,S., Waterston,R. and Wilson,R.  
Apple Functional Genomics grant - NSF 0321702  
Unpublished (2004)

Contact: Schuyler S. Korban  
Apple Functional Genomics grant - NSF 0321702  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Library materials provided by: Schuyler S. Korban Library  
constructed by: K. Gasic Library sequenced by: Washington  
University Genome Sequencing Center  
WashU EST name: aam06d09.y1  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. .367  
/organism="Malus x domestica"  
/mol\_type="mRNA"  
/cultivar="GoldRush"  
/db\_xref="taxon:3750"  
/clone="Mdftr3109hl7"  
/tissue\_type="Fruit"  
/lab\_host="DH10B ampicillin resistant"  
/clone\_lib="Mdftr"  
/note="Vector: pBluescript II SK (+); Site 1: NotI;  
Site 2: EcoRII; Total RNA was extracted separately from  
each stage (young fruitlet (<1cm), young fruitlet (1 cm  
dia.), young fruitlet (12cm dia.), maturing fruit I,  
maturing fruit II, mature fruit), using the 'pine tree'  
method. Poly(A)+mRNA was isolated twice from total RNA  
from each stage using the Oligotex Direct mRNA kit  
(Qiagen). mRNA was reverse transcribed into double  
stranded cDNA using a modified oligo18(dT) primer with an  
identifying tag sequence (see table below). cDNA's from  
different stages were pooled in equal amounts before  
adaptor ligation. Tag identification when sequencing from  
5' end: Stage 1 (young fruitlet) insert 18(A)TCGTG; Stage  
2 (young fruitlet 1cm dia) insert 18(A)TCGTG; Stage 3  
(young fruitlet 12cm dia) insert 18(A)TCGTG; Stage 4  
(maturing fruit I) insert 18(A)TCGA; Stage 5 (maturing  
fruit II) insert 18(A)TCGA; Stage 6 (mature fruit) insert  
18(A)TCGTG; Tag identification when sequencing from 3',  
end: Stage 1 (young fruitlet) CAGCA18(T) insert; Stage 2  
(young fruitlet 1cm dia) CAGCA18(T) insert; Stage 3 (young  
fruitlet 12cm dia) ACCGA18(T) insert; Stage 4 (maturing  
fruit I) TCGCA18(T) insert; Stage 5 (maturing fruit II)  
TCGCA18(T) insert; Stage 6 (mature fruit) ACCGA18(T)  
insert. Double stranded cDNAs were size selected (more  
than 450 bp), adaptored with EcoRI adaptors at both ends  
and then digested with NotI. The cDNAs were then  
directionally cloned into EcoRI-NotI digested pBS II SK(+)  
phagemid vector(Stratagene). Identification of adaptors  
and tags in 5'-end sequenced clones:  
(Vector) . . .TAAGCTT(End Vector)(Start  
EcoRI adaptor)GATATCGAATTCATTGTTGGG (End  
EcoRI\_adaptor)(Start Insert). AAAAAAAAAAAAAAAAAA(End  
Insert) (Start Tag)TCGA(End Tag) (Start  
NotI site/Vector)GGCGCCACCGCG. . . The total number of  
white colony forming units (cfu) in the primary library  
before amplification was 2.1x10<sup>6</sup> cfu (colony forming  
units). The background of empty clones was less than 1%.  
Inserts ranged from 0.5kb to 4 kb, as determined by PCR.  
Purified plasmid DNA from the primary library was  
converted to single-stranded circles and used as a  
template for PCR amplification using the T7 and T3 priming  
sites flanking the cloned cDNA inserts. The purified PCR  
products, representing the entire cloned cDNA population,  
were used as a driver for normalization. Hybridization  
between the single-stranded library and the PCR products  
was carried out for 44 hours at 30C. Unhybridized  
single-stranded DNA circles were separated from hybridized  
DNA rendered partially double-stranded and electroporated  
into DH10B cells to generate the normalized library. The  
total number of clones with insert was 5.6x10<sup>6</sup> cfu.  
Background of empty clones was less than 1%."

FEATURES  
source

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 20:43:21 ; Search time 233.708 Seconds  
(without alignments)  
712.930 Million cell updates/sec

Title: US-10-716-005-3  
Perfect score: 25  
Sequence: 1 caaattaagaagactattcgtgcaa 25

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues 9993994  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : N Geneseq\_21.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*  
14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	25	14	ADZ75863
2	25	100.0	1731	6	ABN68276 Streptoco
3	25	100.0	1734	13	ADV85417 Streptoco
4	25	100.0	29072	13	ADV87723 Streptoco
5	25	100.0	29072	13	ADV78976 Streptoco
6	25	100.0	110000	6	Continuation (8 of
7	25	100.0	110000	13	Continuation (9 of
8	20.2	80.8	525	8	ACA46938 Prokaryot
9	20.2	80.8	1731	8	ACA48261 Prokaryot
10	18.8	75.2	1417	13	ADX45334 Plant ful
11	18.2	72.8	600	9	ADA31807 DNA encod
12	17.8	71.2	1344	8	ACA22236 Prokaryot
13	17.8	71.2	1344	13	ADT41947 Bacterial
14	17.6	70.4	427	3	AAC94729 Cat flea
15	17.6	70.4	489	12	ADM37088 Caenorhab
16	17.6	70.4	512	2	AAT96059 E. coli c
17	17.6	70.4	1173	6	ABN68462 Streptoco
18	17.6	70.4	1431	13	ADT18471 Plant CDN
19	17.6	70.4	2726	13	ADO84764 Plant ful

20	17.6	70.4	2947	4	AAH54441	Aah54441 S. epider
21	17.6	70.4	3285	4	AAH53880	Aah53880 S. epider
22	17.6	70.4	3730	4	AAH54135	Aah54135 S. epider
23	17.6	70.4	10892	4	ABL05132	Abi05132 Drosophila
24	17.6	70.4	19922	8	ACA46526	ACA46526 Prokaryot
25	17.6	70.4	30549	6	ABN90859	Abn90859 Staphyloc
26	17.6	70.4	30612	13	ADS01031	Adso1031 Staphyloc
27	17.6	70.4	49380	4	ABLI1838	Abli1838 Drosophila
28	17.6	70.4	110000	2	AAV21209_15	Continuation (16 of
29	17.6	70.4	110000	13	ABD32968_2	Continuation (3 of
30	17.6	70.4	220224	11	ACN44702	Acn44702 Human gen
31	17.6	70.4	277616	13	ACN432602	Abd32602 Human NFK
32	17.6	69.6	24888	9	ADA02624	Ada02624 Human NFK
33	17.4	69.6	24888	10	ADB72362	Adb72362 Human NFK
34	17.4	69.6	24888	10	ADG59872	Adg59872 Human NFK
35	17.4	69.6	50000	10	ADC56843	Adc56843 Human IKB
36	17.4	69.6	53106	9	ADA03044	Ada03044 Human BAT
37	17.4	69.6	53106	9	ADA66328	Ada66328 Human BAT
38	17.4	69.6	53106	10	ADL27122	Adl27122 Human gen
39	17.4	69.6	53106	11	ADL27122	Adl27122 Human gen
40	17.4	69.6	81800	6	ABK84756	Abk84756 Human CDN
41	17.4	69.6	349980	5	AAH41224	Aah41224 Pyrococcus
42	17.2	68.8	490	3	AAK28412	Aac28412 Human sec
43	17.2	68.8	584	8	ABZ51224	Abz51224 Aspergill
44	17.2	68.8	2876	10	ADF37976	Adf37976 Synchroni
45	17.2	68.8	9398	14	AEA61765	Aea61765 Streptoco
46	17.2	68.8	12264	4	ABA88966	Abas88966 Escherich
47	17.2	68.8	23654	6	ABF78844	Abf78844 E. coli C
48	17.2	68.8	23654	10	ADH80411	Adh80411 Escherich
49	17.2	68.8	32132	4	AAL35943	Aal35943 Human mus
50	17.2	68.8	32132	4	ABX58931	Abx58931 cDNA enco
51	17.2	68.8	32132	8	ABX58931	Abx58931 Human mus
52	17.2	68.8	32132	12	ADJ29681	Adj29681 Human mus
53	17.2	68.8	32132	11	ACN43998_6	Continuation (7 of
54	17.2	68.8	58181	8	ABZ74619	Abz74619 Secreted
55	17.2	68.8	58181	10	ADC21010	Adc21010 Human sec
56	17.2	68.8	58181	10	ABZ68140	Abz68140 Human sec
57	17.2	68.8	101636	14	ADZ12890	Adz12890 Murine ca
58	17.2	68.8	110000	2	AAV21209_10	Continuation (11 of
59	17.2	68.8	110000	11	ACN43998_5	Continuation (6 of
60	17.2	68.8	133632	11	ACN45054	Acn45054 Human gen

ALIGNMENTS

RESULT 1  
ADZ75863  
ID ADZ75863 standard; DNA; 25 BP.

XX AC ADZ75863;

XX DT 28-JUL-2005 (first entry)

XX DE Group B streptococcus phosphotransferase (pts) probe, SEQ ID NO:3.

XX KW Microorganism detection; fluorescence; diagnosis;  
streptococcus infection; infection; gynecology and obstetrics;  
phosphotransferase; probe; ss.

XX OS Streptococcus sp. 'group B'.

XX PH Key Location/Qualifiers  
FT modified\_base 25  
FT /mod\_base a  
FT /note= "3", labeled with fluorescent donor fluorescein"

XX PN US2005106578-A1.

XX PD 19-MAY-2005.

XX PF 18-NOV-2003; 2003US-00716005.

XX 18-NOV-2003; 2003US-00716005.  
 XX (UHLJ/) UHL J R.  
 XX (COCK/) COCKERILL F R.  
 XX (AICH/) AICHINGER C.  
 XX (REIS/) REISER A.  
 XX Uhl JR, Cockerill FR, Aichinger C, Reiser A;  
 XX WPI: 2005-371550/38.  
 XX  
 XX Detecting group B streptococcus, comprises amplifying a sample with pts  
 XX primers, hybridizing the sample with fluorescently labeled pts probes,  
 XX and detecting the presence of fluorescence resonance energy transfer.  
 XX  
 XX Claim 3; SEQ ID NO 3; 13pp; English.  
 XX  
 XX The invention relates to a real-time PCR-based method of detecting the  
 XX presence or absence of group B streptococcus (GBS) bacterial pathogens in  
 XX a biological sample from an individual. The method comprises  
 XX amplification of a conserved region of the phosphotransferase (pts) gene  
 XX of GBS (especially using PCR primers ADZ75861-ADZ75862), detection of the  
 XX amplification product with a pair of fluorescently labeled  
 XX phosphotransferase probes (especially ADZ75863-ADZ75864), and detection  
 XX of the presence or absence of fluorescence resonance energy transfer  
 XX (FRET), where presence of FRET is indicative of the presence of GBS in  
 XX the sample. GBS infection is a leading cause of neonatal morbidity and  
 XX mortality, with infection occurring during childbirth. Currently, it is  
 XX recommended that women are screened for GBS during week 35-37 of  
 XX gestation by a culture-based method which may take up to 72 hours for a  
 XX result. However, many women first present at healthcare facilities at the  
 XX time of labor, and in addition, GBS infection can be transient, so that a  
 XX woman free of GBS at the time of screening may not be free of GBS when  
 XX she is due to give birth. The method of the invention provides a real-  
 XX time assay for the detection of group B streptococcus in a sample, is  
 XX more sensitive and specific than prior art non-culture based methods, and  
 XX can thus be implemented for routine diagnosis of the presence of group B  
 XX streptococcus. Sequences ADZ75863-ADZ75864 represent fluorescently  
 XX labeled GBS phosphotransferase probes specifically claimed for use in the  
 XX method of the invention.  
 XX  
 XX Sequence 25 BP; 11 A; 4 C; 4 G; 6 T; 0 U; 0 Other;  
 SQ Query Match 100.0%; Score 25; DB 14; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 0.13; Length 25;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAATTAAGAGACTATTTCGTGCAA 25  
 DB 1 CAAATTAAGAGACTATTTCGTGCAA 25  
 RESULT 2  
 ABN68276  
 ID ABN68276 standard; DNA; 1731 BP.  
 XX AC  
 XX ABN68276;  
 XX  
 XX 01-JUL-2002 (first entry)  
 XX  
 XX Streptococcus polynucleotide SEQ ID NO 4465.  
 DE  
 XX Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;  
 XX group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
 XX  
 XX Streptococcus agalactiae.  
 OS  
 XX WO200234771-A2.  
 PN  
 XX 02-MAY-2002.  
 PD  
 XX

PF 29-OCT-2001; 2001WO-GB004789.  
 XX  
 XX 27-OCT-2000; 2000GB-00026333.  
 PR 24-NOV-2000; 2000GB-00028727.  
 PR 07-MAR-2001; 2001GB-00005640.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Telford J, Masignani V, Margarit Y Rosl, Grandi G, Fraser C;  
 PI Tettelin H;  
 XX  
 XX WPI: 2002-352536/38.  
 DR P-PSDB; ABP27645.  
 DR  
 XX New Streptococcus protein for the treatment or prevention of infection or  
 PT disease caused by Streptococcus bacteria, such as meningitis, and for  
 PT detecting a compound that binds to the protein.  
 PT  
 XX Claim 7; Page 3607; 4525pp; English.  
 PS  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and antiinflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins  
 XX  
 XX Sequence 1731 BP; 541 A; 323 C; 373 G; 494 T; 0 U; 0 Other;  
 SQ Query Match 100.0%; Score 25; DB 6; Length 1731;  
 Best Local Similarity 100.0%; Pred. No. 0.2; Length 1731;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAATTAAGAGACTATTTCGTGCAA 25  
 DB 265 CAAATTAAGAGACTATTTCGTGCAA 289  
 RESULT 3  
 ADV85417  
 ID ADV85417 standard; DNA; 1734 BP.  
 XX AC  
 XX ADV85417;  
 XX  
 XX 24-FEB-2005 (first entry)  
 XX  
 XX Streptococcus agalactiae DNA sequence, SEQ ID 6558.  
 DE  
 XX Antibacterial; vaccine; bacterial infection; ds.  
 KW  
 XX Streptococcus agalactiae.  
 OS  
 XX WO200292818-A2.  
 PN  
 XX 21-NOV-2002.  
 PD  
 XX 26-APR-2002; 2002WO-IB003059.  
 PF  
 XX 26-APR-2001; 2001FR-00005642.  
 PR  
 XX (INSP ) INST PASTEUR.  
 PA

PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.  
 DR  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 XX Claim 4; SEQ ID NO 6558; 439pp; French.  
 PS  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (I) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (I) or (II) are  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.  
 XX  
 SQ Sequence 1734 BP; 540 A; 323 C; 376 G; 495 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 25; DB 13; Length 1734;  
 Best Local Similarity 100.0%; Pred. No. 0.2;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAATTAAAGAGACTATTCTGCGAA 25  
 DB 265 CAAATTAAAGAGACTATTCTGCGAA 289  
 RESULT 4  
 ID ADV87723 standard; DNA; 29072 BP.  
 AC ADV87723;  
 XX 24-FEB-2005 (first entry)  
 DT Streptococcus agalactiae DNA sequence, SEQ ID 117.  
 XX Antibacterial; Vaccine; bacterial infection; ds.  
 XX Streptococcus agalactiae.  
 XX FR2824074-A1.  
 XX 31-OCT-2002.  
 XX 26-APR-2001; 2001FR-00005642.  
 XX 26-APR-2001; 2001FR-00005642.  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 DR WPI; 2004-101891/11.  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus

PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX Claim 1; SEQ ID NO 117; 2687pp; French.  
 PS  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
 CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
 CC agalactiae involved in the synthesis of amino acids, cell membranes,  
 CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
 CC phospholipid metabolism, nucleotide metabolism including purines,  
 CC pyrimidines and/or nucleosides, regulatory functions, replication,  
 CC transcription, translation, protein transport, adaptation to atypical  
 CC conditions, sensitivity to medicines and/or analogues, functions related  
 CC to transposons, biosynthesis of cofactors, prosthetic groups and  
 CC transporters, cell membrane proteins and cellular machinery. (I) are  
 CC useful for the detection and/or amplification of nucleic acids.  
 CC Pharmaceutical composition comprising (I) or (II) are useful for  
 CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
 CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
 CC contains 6617 sequence whereas the present patent only contains 2344  
 CC sequences.  
 XX  
 SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 25; DB 13; Length 29072;  
 Best Local Similarity 100.0%; Pred. No. 0.26;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAATTAAAGAGACTATTCTGCGAA 25  
 DB 12160 CAAATTAAAGAGACTATTCTGCGAA 12184  
 RESULT 5  
 ID ADV78976 standard; DNA; 29072 BP.  
 XX ADV78976;  
 XX 24-FEB-2005 (first entry)  
 DT Streptococcus agalactiae DNA sequence, SEQ ID 117.  
 XX Antibacterial; vaccine; bacterial infection; ds.  
 XX Streptococcus agalactiae.  
 XX WO200292818-A2.  
 XX 21-NOV-2002.  
 XX 26-APR-2002; 2002WO-IB003059.  
 XX 26-APR-2001; 2001FR-00005642.  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX Claim 1; SEQ ID NO 117; 439pp; French.  
 PS  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The



CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC cellular machinery. (i) are useful for the detection and/or amplification  
 CC of nucleic acids. Pharmaceutical composition comprising (i) or (ii) are  
 CC useful for treatment of a bacterial *S. agalactiae* infection. The complete  
 CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.

SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 25; DB 13; Length 29072;  
 Best Local Similarity 100.0%; Pred. No. 0.26; Mismatches 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAAATTAAAGAGACTATTCTGCGAA 25

Db 12160 CAAATTAAAGAGACTATTCTGCGAA 12184

# RESULT 6

ABN71527\_07  
 Continuation (8 of 22) of ABN71527 from base 700001 (*Streptococcus* polynucleotide SEQ ID  
 WP Sequence split into 22 fragments LOCUS ABN71527 Accession Abn71527

Fragment Name	Begin	End
WP ABN71527_00	1	110000
WP ABN71527_01	100001	210000
WP ABN71527_02	200001	310000
WP ABN71527_03	300001	410000
WP ABN71527_04	400001	510000
WP ABN71527_05	500001	610000
WP ABN71527_06	600001	710000
WP ABN71527_07	700001	810000
WP ABN71527_08	800001	910000
WP ABN71527_09	900001	1010000
WP ABN71527_10	1000001	1110000
WP ABN71527_11	1100001	1210000
WP ABN71527_12	1200001	1310000
WP ABN71527_13	1300001	1410000
WP ABN71527_14	1400001	1510000
WP ABN71527_15	1500001	1610000
WP ABN71527_16	1600001	1710000
WP ABN71527_17	1700001	1810000
WP ABN71527_18	1800001	1910000
WP ABN71527_19	1900001	2010000
WP ABN71527_20	2000001	2110000
WP ABN71527_21	2100001	2155561

Query Match 100.0%; Score 25; DB 6; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 0.29; Mismatches 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAAATTAAAGAGACTATTCTGCGAA 25

Db 91030 CAAATTAAAGAGACTATTCTGCGAA 91054

# RESULT 7

ADV81204\_08  
 Continuation (9 of 23) of ADV81204 from base 800001 (*Streptococcus agalactiae* complete g  
 WP Sequence split into 23 fragments LOCUS ADV81204 Accession Adv81204

Fragment Name	Begin	End
WP ADV81204_00	1	110000
WP ADV81204_01	100001	210000
WP ADV81204_02	200001	310000
WP ADV81204_03	300001	410000
WP ADV81204_04	400001	510000

WP ADV81204_05	500001	610000
WP ADV81204_06	600001	710000
WP ADV81204_07	700001	810000
WP ADV81204_08	800001	910000
WP ADV81204_09	900001	1010000
WP ADV81204_10	1000001	1110000
WP ADV81204_11	1100001	1210000
WP ADV81204_12	1200001	1310000
WP ADV81204_13	1300001	1410000
WP ADV81204_14	1400001	1510000
WP ADV81204_15	1500001	1610000
WP ADV81204_16	1600001	1710000
WP ADV81204_17	1700001	1810000
WP ADV81204_18	1800001	1910000
WP ADV81204_19	1900001	2010000
WP ADV81204_20	2000001	2110000
WP ADV81204_21	2100001	2210000
WP ADV81204_22	2200001	2217924

Query Match 100.0%; Score 25; DB 13; Length 110000;  
 Best Local Similarity 100.0%; Pred. No. 0.29; Mismatches 0; Indels 0; Gaps 0;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CAAATTAAAGAGACTATTCTGCGAA 25

Db 71365 CAAATTAAAGAGACTATTCTGCGAA 71389

# RESULT 8

ACA46938  
 ID ACA46938 standard; DNA; 525 BP.

XX AC ACA46938;

XX DT 19-JUN-2003 (first entry)

XX DE Prokaryotic essential gene #28595.

XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 drug design; gene.

XX OS *Staphylococcus epidermidis*.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 Walli D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX PI WPI: 2003-029926/02.

XX DR P-PSDB; ABU43068.

XX PT New antisense nucleic acids, useful for identifying proteins or screening  
 for homologous nucleic acids required for cellular proliferation to  
 isolate candidate molecules for rational drug discovery programs.

XX PS Claim 14; SEQ ID NO 34808; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of  
 the 6213 antisense sequences given in the specification where expression  
 of the nucleic acid inhibits proliferation of a cell. Also included are:  
 (1) a vector comprising a promoter operably linked to the nucleic acid



CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 525 BP; 214 A; 71 C; 106 G; 134 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 8; Length 525;  
Best Local Similarity 88.0%; Pred. No. 27;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGCAA 25  
||||||| ||||| ||||| ||||| |||||  
DB 28 CAAATTAAAGAGAAATTCGTGCAA 52

RESULT 9  
ACA48261  
ID ACA48261 standard; DNA; 1731 BP.  
XX  
AC ACA48261;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene #29918.  
XX  
KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
OS Streptococcus mutans.  
XX  
PN W0200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
FA (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU44391.

XX New antisense nucleic acids, useful for identifying proteins or screening  
PT cell for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 36131; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1731 BP; 497 A; 320 C; 399 G; 515 T; 0 U; 0 Other;

Query Match 80.8%; Score 20.2; DB 8; Length 1731;  
Best Local Similarity 88.0%; Pred. No. 30;  
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGCAA 25  
||||||| ||||| ||||| ||||| |||||  
DB 265 CAGATTAAAGAGACAATTCGTACAA 289

RESULT 10  
ADX45334  
ID ADX45334 standard; cDNA; 1417 BP.  
XX  
AC ADX45334;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polynucleotide seqid 20074.

XX plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX  
XX Unidentified.  
OS  
XX  
XX US2004034888-A1.  
PN  
XX

```
PD 19-FEB-2004.
XX
XX
PF 28-APR-2003; 2003US-00425114.
XX
XX
PR 06-MAY-1999; 99US-00304517.
PR 05-NOV-2001; 2001US-00985678.
XX
XX
PA (LIUJ/) LIU J.
PA (ZHOU/) ZHOU Y.
PA (KOVA/) KOVALIC D K.
PA (SCRE/) SCREEN S E.
PA (TABA/) TABASKA J E.
PA (CAOY/) CAO Y.
XX
XX
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;
XX
XX
XX WPI; 2004-180133/17.
XX
XX
XX New recombinant DNA construct, useful for improving plant tolerance to
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or
PT pests, for conferring increased resistance to plant disease, or for
PT improving yield.
XX
XX
XX Claim 1; SEQ ID NO 20074; 15pp; English.
XX
XX
XX The invention describes a recombinant DNA construct comprising a
CC polynucleotide consisting of a sequence encoding an amino acid sequence
CC available in electronic form from the US patent office at
CC ftp.segdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide
CC of the invention are also useful in physical arrays of molecules and as
CC plant breeding markers. The recombinant DNA construct is useful for
CC improving plant tolerance to cold, heat, drought, herbicides, extreme
CC osmotic conditions, pathogens or pests, for manipulating growth rate in
CC plant cells by modification of the cell cycle pathway, for conferring
CC increased resistance to plant disease, for producing galactomannan,
CC lignin or plant growth regulators, for increasing the rate of homologous
CC recombination in plants, for improving yield by modification of
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake
CC or by providing improved plant growth and development under at least one
CC stress condition or for modifying seed oil or protein yield and/or
CC content. This sequence represents a plant full length insert
CC polynucleotide that can be used in the recombinant DNA construct of the
CC invention.
XX
XX
XX Sequence 1417 BP; 485 A; 200 C; 247 G; 485 T; 0 U; 0 Other;
SQ
Query Match 75.2%; Score 18.8; DB 13; Length 1417;
Best Local Similarity 90.9%; Pred. NO. 1.3e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 4 ATTAAGAGACTATTCTGTCGA 25
Db 926 ATTAAGAACTATTCTGTGTA 947
RESULT 11
ADA31807
ID ADA31807 standard; DNA; 600 BP.
XX
XX
XX ADA31807;
XX
XX
XX 20-NOV-2003 (first entry)
XX
XX
XX DNA encoding Acinetobacter baumannii protein #3094.
XX
XX
XX ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
XX vaccine; plant biocontrol agent.
XX
XX
XX Acinetobacter baumannii.
XX
XX
XX US6562958-B1.
XX
XX
XX 13-MAY-2003.
XX
PD
19-FEB-2004.
XX
XX
PF 04-JUN-1999; 99US-00328352.
XX
XX
XX 09-JUN-1998; 98US-0088701P.
XX
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX
XX Breton G, Bush D;
XX
XX
XX WPI; 2003-576092/54.
XX
XX
XX P-PSDB; ADA35933.
XX
XX
XX New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
PT for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.
XX
XX
XX Example; SEQ ID NO 3094; 328pp; English.
XX
XX
XX The invention relates to isolated Acinetobacter baumannii nucleic acids.
CC The A. baumannii nucleic acids and polypeptides are useful as reagents
CC for diagnosing a bacterial disease, as components of antibacterial
CC vaccines, as targets for antibacterial drugs, to detect the presence of
CC A. baumannii and other Acinetobacter species in a sample, in screening
CC compounds for the ability to interfere with the A. baumannii life cycle
CC or to inhibit A. baumannii infection, and as biocontrol agents for
CC plants. The present sequence represents DNA encoding an A. baumannii
CC protein.
XX
XX
XX Sequence 600 BP; 178 A; 104 C; 120 G; 198 T; 0 U; 0 Other;
SQ
Query Match 72.8%; Score 18.2; DB 9; Length 600;
Best Local Similarity 87.0%; Pred. NO. 2.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAATTAAAGAGACTATTCTGTCGA 24
Db 83 AAATTGAAGAACTATTCTGTA 105
RESULT 12
ACA22236
ID ACA22236 standard; DNA; 1344 BP.
XX
XX
XX ACA22236;
XX
XX
XX 19-JUN-2003 (first entry)
XX
XX
XX Prokaryotic essential gene #3893.
XX
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
XX drug design; gene.
XX
XX
XX Bacillus anthracis.
XX
XX
XX WO200277183-A2.
XX
XX
XX 03-OCT-2002.
XX
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX
XX 25-OCT-2001; 2001US-0342923P.
XX
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX
XX 06-MAR-2002; 2002US-0362699P.
XX
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX
XX Wang L, Zamudio C, Malone C, Haseibeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX
XX
XX
```

DR P-PSDB; ABU18366.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX  
PS Claim 14; SEQ ID NO 10106; 1765pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1344 BP; 457 A; 210 C; 302 G; 375 T; 0 U; 0 Other;  
  
Query Match 71.2%; Score 17.8; DB 8; Length 1344;  
Best Local Similarity 90.5%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 AAATTAAGAGACTATTTCGTG 22  
Db 1166 AAATTAAGAGAAATTTTCGTG 1186  
|||||  
RESULT 13  
ADT41947  
ID ADT41947 standard; cDNA; 1344 BP.  
XX  
AC ADT41947;  
XX  
XX 02-DEC-2004 (first entry)  
XX  
XX Bacterial polynucleotide #16698.  
XX  
KW Recombinant DNA construct; transformed plant; improved plant property;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;  
KW pathogen tolerance; pest tolerance; plant disease resistance;  
KW cell cycle pathway modification; plant growth regulator;  
KW homologous recombination; seed oil yield; protein yield; carbohydrate;  
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
KW bacterial polynucleotide; gene; ss.  
XX  
XX Bacteria.  
XX  
XX OS  
XX PN US2003233675-A1.  
XX

PD 18-DEC-2003.  
XX  
XX 20-FEB-2003; 2003US-00369493.  
XX  
PR 21-FEB-2002; 2002US-0360039P.  
XX  
PA (CAOY/) CAO Y.  
PA (HINK/) HINKLE G J.  
PA (SLAT/) SLATER S C.  
PA (CHEN/) CHEN X.  
PA (GOLD/) GOLDMAN B S.  
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX WPI; 2004-061375/06.  
XX  
XX New recombinant DNA construct comprising a promoter positioned to provide  
PT for expression of a polynucleotide encoding a polypeptide from a  
PT microbial source, useful for producing plants with improved properties.  
XX  
XX Claim 1; SEQ ID NO 40385; 122pp; English.  
XX  
CC The invention relates to a recombinant DNA construct comprising a  
CC promoter functional in a plant cell, where the promoter is positioned to  
CC provide for expression of a polynucleotide encoding a polypeptide from a  
CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plants with  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at seqdata.uspto.gov/sequence.html.  
XX  
SQ Sequence 1344 BP; 451 A; 213 C; 303 G; 377 T; 0 U; 0 Other;  
  
Query Match 71.2%; Score 17.8; DB 13; Length 1344;  
Best Local Similarity 90.5%; Pred. No. 3.7e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
QY 2 AAATTAAGAGACTATTTCGTG 22  
Db 1166 AAATTAAGAGAAATTTTCGTG 1186  
|||||  
RESULT 14  
AAC94729/c  
ID AAC94729 standard; cDNA; 427 BP.  
XX  
XX AAC94729;  
XX  
XX 19-FEB-2001 (first entry)  
XX  
XX Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:1224.  
XX  
XX Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;  
KW flea infestation; vaccine; antiparasitic; therapeutic target; diagnosis;  
KW detection; ss.  
XX  
XX Ctenocephalides felis.  
OS

```
XX WO200061621-A2.
XX 19-OCT-2000.
XX 07-APR-2000; 2000WO-US009437.
XX 09-APR-1999; 99US-0128704P.
XX (HESK-) HESKA CORP.
XX Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
XX WPI; 2000-656323/63.
XX Flea Malpighian tubule and head and nerve cord tissue derived nucleic
XX acids useful for the prevention, diagnosis and treatment of flea
XX infestations.
XX Claim 26; Page 640; 964pp; English.
XX The invention relates to novel cat flea (Ctenocephalides felis) nucleic
XX acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
XX or head and nerve cord (HNC) tissue. The invention also relates to the
XX encoded proteins. The invention additionally encompasses expression
XX constructs, recombinant viruses and recombinant cells comprising the
XX nucleic acids of the invention, recombinant production of the proteins,
XX antibodies against the proteins, a method of identifying inhibitors of
XX the proteins, and compositions comprising the inhibitors for
XX administration to an animal. The nucleic acids, and the proteins they
XX encode may be used in the prevention, treatment and diagnosis of diseases
XX associated with flea infestations. For example, the nucleic acids may be
XX used to produce an HMT or HNC protein according to standard recombinant
XX DNA methodology by inserting the nucleic acids into a host cell and
XX culturing the cell to express the protein. The HMT and HNC nucleic acids
XX may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
XX and quantitate the presence of cat flea or other homologous nucleic acid
XX sequences in samples. They may also be used to study the expression and
XX function of the proteins and their role in metabolism. The HMT and HNC
XX proteins may be used as antigens in the production of specific
XX antibodies, and in assays to identify modulators (agonists and
XX antagonists) of HMT and/or HNC protein expression and activity. The anti-
XX HMT/HNC protein antibodies and antagonists may also be used to
XX downregulate protein expression and activity. The antibodies may also be
XX used as diagnostic agents for detecting the presence of flea polypeptides
XX in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
XX present sequence represents a cat flea HMT cDNA of the invention
XX
XX Sequence 427 BP; 180 A; 57 C; 38 G; 152 T; 0 U; 0 Other;
SQ Query Match 70.4%; Score 17.6; DB 3; Length 427;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 AAATTAAGAGAGACTATTCGTGCA 25
Db 359 AAATTTAAGAAACTATTCATGTA 336
RESULT 15
ADM37088
ID ADM37088 standard; cDNA; 489 BP.
XX AC ADM37088;
XX
XX 03-JUN-2004 (first entry)
XX
XX Caenorhabditis elegans STARS encoding cDNA SEQ ID NO:7.
XX striated muscle activator of Rho signalin; STARS;
XX muscle-specific actin-binding protein; cardiant; gene therapy;
XX cardiovascular disease; cardiac hypertrophy; dilated cardiomyopathy;
XX myocardial infarct; heart failure; drug screening; gene; ss.
```

```
XX OS Caenorhabditis elegans.
XX Key Location/Qualifiers
XX CDS 1..489
XX /*tag= a
XX WO2004018629-A2.
XX 04-MAR-2004.
XX 20-AUG-2003; 2003WO-US026191.
XX 20-AUG-2002; 2002US-0404706P.
XX (TEXA) UNIV TEXAS SYSTEM.
XX Olson E, Arai A;
XX WPI; 2004-226822/21.
XX P-PSDB; ADM37089.
XX New striated muscle activator of Rho signalin polypeptides and nucleic
XX acid molecules useful for preventing and treating cardiovascular
XX diseases, e.g. cardiac hypertrophy, myocardial infarct or heart failure,
XX or in drug screening.
XX Claim 3; SEQ ID NO 7; 127pp; English.
XX The present sequence encodes a Caenorhabditis elegans striated muscle
XX activator of Rho signalin (STARS) protein, which is a muscle-specific
XX actin-binding protein. Also described: (1) a polypeptide comprising any
XX of the 5 STARS sequences of 162-399 amino acids (see SEQ ID NO:2, 4, 6, 8
XX or 10), or a peptide comprising 10 contiguous amino acids of any of the
XX above amino acid sequences; (2) an expression construct comprising a
XX polynucleotide encoding a STARS polypeptide operably linked to a
XX regulatory sequence; (3) methods of screening for modulators of STARS
XX expression or STARS actin-binding activity; (4) a method of screening for
XX an inhibitor of STARS-induced transcription; (5) a method of producing a
XX STARS polypeptide in a cell; (6) a non-human transgenic animal comprising
XX a selectable or screenable marker protein under the control of a STARS
XX promoter, or comprising a STARS encoding nucleic acid under the control
XX of an inducible or a constitutive promoter; (7) a method of inhibiting
XX STARS activity; (8) methods of treating or inhibiting progression of
XX cardiac hypertrophy, dilated cardiomyopathy, myocardial infarct or heart
XX failure; (9) a method of preventing cardiac hypertrophy and dilated
XX cardiomyopathy; (10) methods of increasing exercise tolerance, reducing
XX hospitalisation, improving the quality of life or decreasing mortality
XX and morbidity in a subject with heart failure or cardiac hypertrophy;
XX (11) methods of producing a modulator of STARS expression or STARS actin
XX binding activity; (12) a modulator of STARS expression identified by the
XX method in (3); (13) an antibody that binds immunologically to STARS, or a
XX polyclonal antibody preparation, antibodies of which bind immunologically
XX to STARS; and (14) a hybridoma cell that produces a monoclonal antibody
XX that binds immunologically to STARS. STARS sequences have cardiant
XX activities, and can be used in gene therapy. Compositions and methods of
XX the present invention are useful for preventing and treating
XX cardiovascular diseases, such as cardiac hypertrophy, dilated
XX cardiomyopathy, myocardial infarct or heart failure. They may also be
XX used in drug screening applications.
XX
XX Sequence 489 BP; 186 A; 81 C; 103 G; 119 T; 0 U; 0 Other;
SQ Query Match 70.4%; Score 17.6; DB 12; Length 489;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 CAATTAAGAGAGACTATTCGTGCA 24
Db 439 CAGCTTAAAGAGGCTATTCGAGCA 462
RESULT 16
```

```

AAT96059
ID AAT96059 standard; DNA; 512 BP.
XX
AC AAT96059;
XX
XX
XX 27-MAR-1998 (first entry)
XX
DE E. coli colonisation factor antigen CPA1 DNA.
XX
XX Bacterial colonisation; colonisation factor antigen; CPA1;
KW enterotoxigenic Escherichia coli; vaccine; diagnosis; research; ds.
KW
XX Escherichia coli.
OS
XX
XX Key Location/Qualifiers
FH 1..512
FT /*tag= a
FT
XX
XX US5698416-A.
XX
XX 16-DEC-1997.
XX
XX 02-JUN-1995; 95US-00460739.
PF
XX
XX 02-JUN-1995; 95US-00460739.
PR
XX
XX (USSA ) US SEC OF ARMY.
PA
XX
XX Bell BA, Wolf MK, Cassels FJ;
PI
XX WPI; 1998-051486/05.
DR
XX P-PSDB; AAW38341.
DR
XX
XX Production of bacterial colonisation factor protein - by expression under
PT control of heat-inducible promoter.
PT
XX
XX Example 2; Col 15-16; 11pp; English.
XX
XX Production of a protein that affects bacterial colonisation, comprises
CC inoculating a broth containing tryptone and yeast extract with enteric
CC bacteria containing a DNA sequence encoding the protein under the control
CC of a temperature regulated promoter, culturing the bacteria, removing the
CC bacteria from the medium and recovering the protein. The method is used
CC especially for producing the colonisation factor antigen CPA1 of
CC enterotoxigenic E. coli, i.e. the antigen encoded by the present sequence,
CC which may be used in vaccines or for diagnostic or research purposes.
CC Growing the bacteria at low temperature until the late logarithmic phase
CC increases the yield of the protein
XX
XX Sequence 512 BP; 160 A; 99 C; 104 G; 149 T; 0 U; 0 Other;
SQ
Query Match 70.4%; Score 17.6; DB 2; Length 512;
Best Local Similarity 83.3%; Pred. No. 4.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AATTTAAAGAGACTATTTCGTGCAA 25
Db ||||||| ||||||| ||||||| |||||||
5 AATTTAAAGAGACTATTTCGTGCAA 28

RESULT 17
ABN68462
ID ABN68462 standard; DNA; 1173 BP.
XX
AC ABN68462;
XX
XX 01-JUL-2002 (first entry)
XX
XX Streptococcus polynucleotide SEQ ID NO 4837.
DE
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
KW

```

---

```

XX Streptococcus pyogenes.
OS
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
PR
XX 24-NOV-2000; 2000GB-00028727.
PR
XX 07-MAR-2001; 2001GB-00005640.
PR
XX
XX (CHIR-) CHIRON SPA.
PA
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y Rosi, Grandi G, Fraser C;
PI Tettelin H;
PI
XX
XX WPI; 2002-352536/38.
XX
XX P-PSDB; ABP27831.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
PT disease caused by Streptococcus bacteria, such as meningitis, and for
PT detecting a compound that binds to the protein.
XX
XX Claim 7; Page 3646; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins
XX
XX Sequence 1173 BP; 397 A; 198 C; 237 G; 341 T; 0 U; 0 Other;
SQ
Query Match 70.4%; Score 17.6; DB 6; Length 1173;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AATTTAAAGAGACTATTTCGTGCAA 25
Db ||||||| ||||||| ||||||| |||||||
773 AATTTAAAGAGCTAGACGTGCAA 796

RESULT 18
ADT18471/c
ID ADT18471 standard; cDNA; 1431 BP.
XX
XX
XX ADT18471;
XX
XX 13-JAN-2005 (first entry)
XX
XX Plant cDNA, Seq ID 3797.
XX
XX Plant; ss; gene; transgenic; cold tolerance; growth rate;
KW drought tolerance; disease resistance; galactomannan production;
KW plant growth regulator; heat tolerance; herbicide tolerance;
KW lignin production; extreme osmotic condition tolerance;
KW pathogens resistance; pest resistance; yield improvement; seed oil yield;
KW seed protein yield.

```

XX Viridiplantae.  
OS US2004216190-A1.  
PN 28-OCT-2004.  
XX 18-DEC-2003; 2003US-00739930.  
XX 28-APR-2003; 2003US-00424599.  
PR 28-APR-2003; 2003US-00425115.  
XX (KOVA/) KOVALIC D K.  
PA Kovalic DK;  
PI WPI; 2004-757369/74.  
DR New recombinant DNA constructs useful in the field of biochemistry and  
PT genetics, and in particular for producing transgenic plants with improved  
PT biological characteristics.  
XX  
PS Claim 1; SEQ ID NO 3797; 14pp; English.  
XX  
XX The invention relates a recombinant DNA construct comprising a  
CC polynucleotide having any of 5344 nucleotide sequences (CDNAs SEQ ID NO:  
CC 1-5544) and encoding a polypeptide with any of 5544 amino acid sequences  
CC (SEQ ID NO: 5545-11088). The CDNAs and proteins are from corn, soybean,  
CC Arabidopsis, wheat and tape but the specification does not indicate which  
CC sequences is derived from which organism. Also included is a method of  
CC producing a plant having an improved property, comprising transforming a  
CC plant with a recombinant DNA construct comprising a promoter region  
CC functional in a plant cell operably joined to a polynucleotide encoding a  
CC polypeptide associated with the property, and growing the transformed  
CC plant. The property is selected from improving plant cold tolerance, for  
CC manipulating growth rate in plant cells by modification of the cell cycle  
CC pathway, for improving plant drought tolerance, for providing increased  
CC resistance to plant disease, for galactomannan production, for production  
CC of plant growth regulators, for improving plant heat tolerance, for  
CC improving plant tolerance to herbicides, for increasing the rate of  
CC homologous recombination in plants, for lignin production, for improving  
CC plant tolerance to extreme osmotic conditions, for improving plant  
CC tolerance to pathogens or pests, for yield improvement by modification of  
CC photosynthesis, for modifying seed oil yield and/or content, for  
CC modifying seed protein yield and/or content, for yield improvement by  
CC modification of carbohydrate, nitrogen or phosphorus use and/or uptake  
CC and for yield improvement by providing improved plant growth and  
CC development under at least one stress condition. The polynucleotide may  
CC also encode a plant transcription factor. The methods and compositions of  
CC the present invention are useful in the field of biochemistry and  
CC genetics, in particular for producing transgenic plants with improved  
CC biological characteristics such as increased yield, improved nitrogen  
CC flow, increasing plant tolerance to cold or heat, improving plant  
CC tolerance to extreme osmotic and drought conditions, and improving plant  
CC tolerance to plant pests or pathogens. They can also be used in physical  
CC arrays of molecules, plant breeding markers, computer-based storage and  
CC analysis systems. The present sequence is one of the 5544 plant CDNA  
CC sequences of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20040216190.  
XX  
SQ Sequence 1431 BP; 496 A; 246 C; 283 G; 406 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 13; Length 1431;  
Best Local Similarity 83.3%; Pred. No. 4.6e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 2 AAATTAAAGAGACTATTCGTGCA 25  
|||||  
Db 396 AAATTAAAGAACTACTCTTGAA 373  
|||||

RESULT 19  
ADO84764/C  
ID ADO84764 standard; cDNA; 2726 BP.  
XX  
AC ADO84764;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Plant full length insert polynucleotide seqid 3484.  
XX  
KW plant protectant; plant growth regulant; gene therapy; plant;  
KW recombinant DNA construct; physical array; plant breeding marker;  
KW cold tolerance; heat tolerance; drought tolerance; herbicide tolerance;  
KW extreme osmotic condition; pathogen tolerance; pest tolerance;  
KW growth rate; cell cycle pathway; disease resistance;  
KW galactomannan production; lignin production; plant growth regulator;  
KW yield; plant growth; plant development; seed oil; protein yield;  
KW protein content; gene; ss.  
XX  
OS Unidentified.  
XX  
PN US2004034888-A1.  
XX  
XX 19-FEB-2004.  
XX  
XX 28-APR-2003; 2003US-00425114.  
PF  
XX 06-MAY-1999; 99US-00304517.  
PR  
XX 05-NOV-2001; 2001US-00985678.  
PR  
XX (LIU/) LIU J.  
PA (ZHOU/) ZHOU Y.  
PA (KOVA/) KOVALIC D K.  
PA (SCRE/) SCREEN S E.  
PA (TAB/) TABASKA J E.  
PA (CAO/) CAO Y.  
XX  
PI Liu J, Zhou Y, Kovalic DK, Screen SE, Tabaska JE, Cao Y;  
XX WPI; 2004-180133/17.  
XX  
XX New recombinant DNA construct, useful for improving plant tolerance to  
PT cold, heat, drought, herbicides, extreme osmotic conditions, pathogens or  
PT pests, for conferring increased resistance to plant disease, or for  
PT improving yield.  
XX  
PS Claim 1; SEQ ID NO 3484; 15pp; English.  
XX  
XX The invention describes a recombinant DNA construct comprising a  
CC polynucleotide consisting of a sequence encoding an amino acid sequence  
CC available in electronic form from the US patent office at  
CC ftp.seqdata.uspto.gov/sequence.html?DocID:2004034888. The polynucleotide  
CC of the invention are also useful in physical arrays of molecules and as  
CC plant breeding markers. The recombinant DNA construct is useful for  
CC improving plant tolerance to cold, heat, drought, herbicides, extreme  
CC osmotic conditions, pathogens or pests, for manipulating growth rate in  
CC plant cells by modification of the cell cycle pathway, for conferring  
CC increased resistance to plant disease, for producing galactomannan,  
CC lignin or plant growth regulators, for increasing the rate of homologous  
CC recombination in plants, for improving yield by modification of  
CC photosynthesis or carbohydrate, nitrogen or phosphorus use and/or uptake  
CC or by providing improved plant growth and development under at least one  
CC stress condition or for modifying seed oil or protein yield and/or  
CC content. This sequence represents a plant full length insert  
CC polynucleotide that can be used in the recombinant DNA construct of the  
CC invention.  
XX  
SQ Sequence 2726 BP; 655 A; 586 C; 662 G; 823 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 13; Length 2726;  
Best Local Similarity 83.3%; Pred. No. 4.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
Db 528 CAAATCAGAGACATTCCTGTCGA 505

RESULT 20
AAH54441
ID AAH54441 standard; DNA; 2947 BP.
XX AC AAH54441;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3805.
XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
XX KW endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX PT useful for vaccinating against infections, e.g. endocarditis.
XX PS Claim 8; Page 1441-1442; 2188pp; English.
XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
XX CC and (II) can have antibacterial activity and therefore can be used in
XX CC vaccination. The nucleic acids (I) may be used to produce the S.
XX CC epidermidis polypeptides (II) via the production of vectors containing
XX CC them which are used to produce hosts cells which express the
XX CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX CC used to vaccinate subjects and to raise antibodies against the bacteria.
XX CC The polypeptides may also be used to assay for other inhibitors of their
XX CC activity and therefore identify compounds that may be used for the
XX CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
XX CC represent oligonucleotide sequences and primers which are used in the
XX CC exemplification of the present invention. N.B. The present invention
XX CC specifically claims all the polynucleotide sequences given in the
XX CC sequence listing of the present specification, however the sequence
XX CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
XX CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
XX CC for SEQ ID NO:4455 to 4464
XX SQ Sequence 2947 BP; 1293 A; 537 C; 448 G; 669 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 2947;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
Db 1894 CAAATAAGAGAGATAATAGTGCA 1917

RESULT 21
AAH53880
ID AAH54135 standard; DNA; 3730 BP.
XX AC AAH54135;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:3153.
XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
XX KW endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX PT useful for vaccinating against infections, e.g. endocarditis.
XX PS Claim 8; Page 831-832; 2188pp; English.
XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
XX CC and (II) can have antibacterial activity and therefore can be used in
XX CC vaccination. The nucleic acids (I) may be used to produce the S.
XX CC epidermidis polypeptides (II) via the production of vectors containing
XX CC them which are used to produce hosts cells which express the
XX CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX CC used to vaccinate subjects and to raise antibodies against the bacteria.
XX CC The polypeptides may also be used to assay for other inhibitors of their
XX CC activity and therefore identify compounds that may be used for the
XX CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
XX CC represent oligonucleotide sequences and primers which are used in the
XX CC exemplification of the present invention. N.B. The present invention
XX CC specifically claims all the polynucleotide sequences given in the
XX CC sequence listing of the present specification, however the sequence
XX CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
XX CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
XX CC for SEQ ID NO:4455 to 4464
XX SQ Sequence 3285 BP; 1462 A; 580 C; 498 G; 745 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 3285;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
Db 1447 CAAATAAGAGAGATAATAGTGCA 1470

RESULT 22
AAH54135
ID AAH54135 standard; DNA; 3730 BP.
XX AC AAH54135;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:3153.
XX KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;
XX KW endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US030782.
XX PR 09-NOV-1999; 99US-0164258P.
XX PA (GLAX ) GLAXO GROUP LTD.
XX PI Kimmerly WJ;
XX DR WPI; 2001-316495/33.
XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX PT useful for vaccinating against infections, e.g. endocarditis.
XX PS Claim 8; Page 831-832; 2188pp; English.
XX CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)
XX CC and (II) can have antibacterial activity and therefore can be used in
XX CC vaccination. The nucleic acids (I) may be used to produce the S.
XX CC epidermidis polypeptides (II) via the production of vectors containing
XX CC them which are used to produce hosts cells which express the
XX CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX CC used to vaccinate subjects and to raise antibodies against the bacteria.
XX CC The polypeptides may also be used to assay for other inhibitors of their
XX CC activity and therefore identify compounds that may be used for the
XX CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX CC polynucleotide sequences from the present invention. AAH55091 to AAH55098
XX CC represent oligonucleotide sequences and primers which are used in the
XX CC exemplification of the present invention. N.B. The present invention
XX CC specifically claims all the polynucleotide sequences given in the
XX CC sequence listing of the present specification, however the sequence
XX CC listing only goes up to SEQ ID NO:4454 so even though sequences are given
XX CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present
XX CC for SEQ ID NO:4455 to 4464
XX SQ Sequence 3285 BP; 1462 A; 580 C; 498 G; 745 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 4; Length 3285;
Best Local Similarity 83.3%; Pred. No. 4.9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```



DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3499.  
XX  
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis; vaccination;  
KW endocarditis; ds.  
XX  
XX Staphylococcus epidermidis.  
XX  
PN WO200134809-A2.  
XX  
PD 17-MAY-2001.  
XX  
XX 09-NOV-2000; 2000WO-US030782.  
XX  
XX 09-NOV-1999; 99US-0164258P.  
PR (GLAX ) GLAXO GROUP LTD.  
XX  
XX Kimmerly WJ;  
XX  
XX WPI; 2001-316495/33.  
DR  
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
PT useful for vaccinating against infections, e.g. endocarditis.  
PT  
XX Claim 8; Page 1071-1072; 2188pp; English.  
XX  
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis. (I)  
CC and (II) can have antibacterial activity and therefore can be used in  
CC vaccination. The nucleic acids (I) may be used to produce the S.  
CC epidermidis polypeptides (II) via the production of vectors containing  
CC them which are used to produce hosts cells which express the  
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
CC used to vaccinate subjects and to raise antibodies against the bacteria.  
CC The polypeptides may also be used to assay for other inhibitors of their  
CC activity and therefore identify compounds that may be used for the  
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
CC AAH5090 represent specifically claimed S. epidermidis genomic DNA  
CC polynucleotide sequences from the present invention. AAH5091 to AAH5098  
CC represent oligonucleotide sequences and primers which are used in the  
CC exemplification of the present invention. N.B. The present invention  
CC specifically claims all the polynucleotide sequences given in the  
CC sequence listing of the present specification, however the sequence  
CC listing only goes up to SEQ ID NO:4454 so even though sequences are given  
CC in the disclosure for SEQ ID NO:4465 to 4472, no sequences are present  
CC for SEQ ID NO:4455 to 4464  
XX  
SQ Sequence 3730 BP; 1657 A; 658 C; 560 G; 855 T; 0 U; 0 Other;  
  
Query Match 70.4%; Score 17.6; DB 4; Length 3730;  
Best Local Similarity 83.3%; Pred. No. 5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAGAGACTATTCTGCA 24  
DB 1892 CAAATAAAGAGAAATAGTGCA 1915  
  
RESULT 23  
ABJ05132  
ID ABL05132 standard; cDNA; 10892 BP.  
XX  
AC ABL05132;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 9878.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
KW  
XX Drosophila melanogaster.  
OS  
XX

PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-655860/75.  
DR P-PsDB; ABB61029.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 9878; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABJ16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 10892 BP; 3489 A; 2712 C; 1869 G; 2822 T; 0 U; 0 Other;  
  
Query Match 70.4%; Score 17.6; DB 4; Length 10892;  
Best Local Similarity 83.3%; Pred. No. 5.5e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 2 AAATTAAGAGACTATTCTGCAA 25  
DB 7456 AAATAAATCAGACTATTGTGCAA 7479  
  
RESULT 24  
ACA46526  
ID ACA46526 standard; DNA; 19922 BP.  
XX  
AC ACA46526;  
XX  
XX 19-JUN-2003 (first entry)  
XX  
XX Prokaryotic essential gene #28183.  
XX  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX Staphylococcus epidermidis.  
XX  
XX WO200277183-A2.  
XX  
XX 03-OCT-2002.  
XX  
XX 21-MAR-2002; 2002WO-US009107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX



PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU42656.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX  
 XX Claim 14; SEQ ID NO 34396; 1766pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 19922 BP; 8789 A; 3297 C; 2990 G; 4846 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 17.6; DB 8; Length 19922;  
 Best Local Similarity 83.3%; Pred. No. 5.8e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 CAAATTAAGAGACTATTCGTGCA 24  
 DB 1327 CAAATAAAGAGAAATAATAGTGCA 1350  
 RESULT 25  
 ABN90859  
 ID ABN90859 standard; DNA; 30549 BP.  
 XX  
 AC ABN90859;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:322.  
 XX  
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy; gene; ds.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.

XX 13-AUG-1998; 98US-00134001.  
 XX  
 XX 14-AUG-1997; 97US-0055779P.  
 PR 08-NOV-1997; 97US-0064964P.  
 XX  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 XX Doucette-Stamm LA, Bush D;  
 PI WPI; 2002-381255/41.  
 XX P-PSDB; ABP38314.  
 DR  
 DR Novel isolated nucleic acid encoding a *Staphylococcus epidermis*  
 PT polypeptide, useful for diagnosing and treating bacterial infections.  
 PT  
 XX Disclosure; SEQ ID NO 322; 267pp; English.  
 PS  
 XX ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The *S. epidermidis* sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly *S. epidermidis* infections. The sequences can be used to  
 CC screen for compounds able to interfere with the *S. epidermidis* life cycle  
 CC or inhibit *S. epidermidis* infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX  
 SQ Sequence 30549 BP; 13033 A; 5054 C; 4748 G; 7714 T; 0 U; 0 Other;  
 Query Match 70.4%; Score 17.6; DB 6; Length 30549;  
 Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 OY 1 CAAATTAAGAGACTATTCGTGCA 24  
 DB 12031 CAAATAAAGAGAAATAATAGTGCA 12054  
 RESULT 26  
 ADS01031  
 ID ADS01031 standard; DNA; 30612 BP.  
 XX  
 AC ADS01031;  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX  
 DE Staphylococcus epidermis polynucleotide seqid 326.  
 XX  
 KW antibacterial; vaccine; antisense therapy; *Staphylococcus epidermidis*;  
 KW recombinant expression vector; infection; computer readable medium;  
 KW computer based system; gene; ds.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US2004147734-A1.  
 XX  
 PD 29-JUL-2004.  
 XX  
 XX 01-DEC-2003; 2003US-00724972.  
 PF  
 XX 08-NOV-1997; 97US-0064964P.  
 PR 13-AUG-1998; 98US-00134001.  
 PR 29-NOV-1999; 99US-00450969.  
 XX  
 XX (DOUC/) DOUCETTE-STAMM L.  
 PA (BUSH/) BUSH D.  
 XX  
 XX Doucette-Stamm L, Bush D;  
 PI WPI; 2004-580138/56.  
 XX P-PSDB; ADS04803.  
 DR

XX New isolated polypeptide and encoding nucleic acid derived from  
PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
PT treating an S. epidermidis bacterial infection.  
XX  
XX Claim 5; SEQ ID NO 326; 741pp; English.  
XX  
CC The invention describes an isolated nucleic acid comprising a nucleotide  
CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
CC 1-3772) and encoding an Staphylococcus epidermidis polypeptide with any  
CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
CC given in the specification. Also described are: a recombinant expression  
CC vector; a cell comprising a recombinant expression vector of (1);  
CC producing an S. epidermidis polypeptide; an isolated nucleic acid  
CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
CC vaccine composition for prevention or treatment of an S. epidermidis  
CC infection, comprising a nucleic acid cited above and a carrier; treating  
CC a subject for S. epidermidis infection; a recombinant or substantially  
CC pure preparation of an S. epidermidis polypeptide or its fragment; a  
CC vaccine composition for prevention or treatment of an S. epidermidis  
CC infection; detecting the presence of a Staphylococcus nucleic acid in a  
CC sample; a computer readable medium having recorded in it the nucleotide  
CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
CC system for identifying fragments of the Staphylococcus genome of  
CC commercial importance; a computer based system for identifying fragments  
CC of the Staphylococcus plasmids of commercial importance; identifying  
CC commercially important nucleic acid fragments of the Staphylococcus  
CC genome and/or plasmids; and identifying an expression modulating fragment  
CC of the Staphylococcus genome and/or plasmids. The methods and  
CC compositions of the present invention are useful for the diagnosis,  
CC prevention and/or treatment of an Staphylococcal epidermidis bacterial  
CC infection. This sequence encodes a S. epidermis protein of the invention.  
XX  
SQ Sequence 30612 BP; 13069 A; 5063 C; 4756 G; 7724 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 13; Length 30612;  
Best Local Similarity 83.3%; Pred. No. 6.1e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCGTGCA 24  
||| ||||| ||| |||||  
Db 12031 CAATTAAGAGACTATTCGTGCA 12054

RESULT 27  
ABL11838  
ID ABL11838 standard; cDNA; 49380 BP.  
XX  
AC ABL11838;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 29996.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
XX  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX

DR P-PSDB; ABB67735.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 29996; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA  
CC sequences (ABU01840-ABU16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 49380 BP; 14230 A; 10576 C; 10238 G; 14336 T; 0 U; 0 Other;  
Query Match 70.4%; Score 17.6; DB 4; Length 49380;  
Best Local Similarity 83.3%; Pred. No. 6.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCGTGCA 24  
||| ||||| ||| |||||  
Db 7624 CAATTAAGAGACTATTCGTGCA 7647

RESULT 28  
AAV21209\_15  
Continuation (16 of 17) of AAV21209 from base 1500001 (Methanococcus jannaschii circular  
WP Sequence split into 17 fragments LOCUS AAV21209 Accession AAV21209  
WP Fragment Name Begin End  
WP AAV21209\_00 1 110000  
WP AAV21209\_01 100001 210000  
WP AAV21209\_02 200001 310000  
WP AAV21209\_03 300001 410000  
WP AAV21209\_04 400001 510000  
WP AAV21209\_05 500001 610000  
WP AAV21209\_06 600001 710000  
WP AAV21209\_07 700001 810000  
WP AAV21209\_08 800001 910000  
WP AAV21209\_09 900001 1010000  
WP AAV21209\_10 1000001 1110000  
WP AAV21209\_11 1100001 1210000  
WP AAV21209\_12 1200001 1310000  
WP AAV21209\_13 1300001 1410000  
WP AAV21209\_14 1400001 1510000  
WP AAV21209\_15 1500001 1610000  
WP AAV21209\_16 1600001 1664976

Query Match 70.4%; Score 17.6; DB 2; Length 110000;  
Best Local Similarity 83.3%; Pred. No. 6.8e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCA 25  
||| ||||| ||| |||||  
Db 22836 AAATTAAGAGAAATTAATGCA 22859

RESULT 29  
ABD32968\_2  
Continuation (3 of 8) of ABD32968 from base 200001 (Human cancer-associated genomic DNA  
WP Sequence split into 8 fragments LOCUS ABD32968 Accession ABD32968  
WP Fragment Name Begin End  
WP ABD32968\_0 1 110000  
WP ABD32968\_1 100001 210000  
WP ABD32968\_2 200001 310000  
WP ABD32968\_3 300001 410000  
WP ABD32968\_4 400001 510000  
WP ABD32968\_5 500001 610000

```
WP ABD32968.6 600001 710000
WP ABD32968.7 700001 779603

Query Match Score 17.6; DB 13; Length 110000;
Best Local Similarity 83.3%; Pred. No. 6.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25
Db 60109 AAATTAAGAGACTAATAATGCAA 60132

RESULT 30
ACN44702/C
ID ACN44702 standard; DNA; 220224 BP.
XX
AC AC
XX
DT 18-NOV-2004 (first entry)
XX
DE Human genomic sequence hCG1788599.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS Homo sapiens.
XX
FN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
XX WPI; 2003-328604/31.
XX
DR Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
PT
XX Claim 1; SEQ ID NO 1282; Opp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
XX Sequence 220224 BP; 61100 A; 44820 C; 47808 G; 64675 T; 0 U; 1821 Other;
SQ

Query Match 70.4%; Score 17.6; DB 11; Length 220224;
Best Local Similarity 83.3%; Pred. No. 7.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25
Db 150331 AAATTAAGAGAGATTGGTTCAA 150308

RESULT 31
ABD32602
```

ABD32602 standard; DNA; 277616 BP.

ABD32602;

18-NOV-2004 (first entry)

Human cancer-associated genomic DNA HD7-221.

Human; ds; cancer-associated protein; gene; cytostatic; cancer;

leukaemia; lymphoma; CAP.

Homo sapiens.

WO2004074320-A2.

02-SEP-2004.

17-FEB-2004; 2004WO-US004730.

14-FEB-2003; 2003US-00367094.

14-MAR-2003; 2003US-00388838.

15-APR-2003; 2003US-00417375.

13-JUN-2003; 2003US-00461862.

15-SEP-2003; 2003US-00683431.

15-DEC-2003; 2003US-00737318.

(SAGR-) SAGRES DISCOVERY INC.

Morris DW, Morris DW, Malandro MS;

WPI; 2004-652914/63.

New isolated cancer-associated polynucleotides and polypeptides useful for diagnosing, preventing or treating cancers, especially lymphoma and leukemia, or in screening for agents that modulate cancer.

claim 16; seqid 109; 310pp; English.

The invention relates to an isolated nucleic acid comprising at least 10 contiguous nucleotides of any of the 233 polynucleotide sequences given in the specification, or its complement. The nucleic acids encode cancer-associated proteins. Also included are an expression vector comprising the isolated nucleic acid cited above, a host cell comprising the above recombinant nucleic acid or expression vector, a microarray for detecting a cancer-associated (CA) nucleic acid comprising at least one probe comprising at least 10 contiguous nucleotides of any of the above-mentioned nucleotide sequences, an isolated polypeptide (encoded within an open reading frame of a CA sequence selected from any of the 95 polynucleotide sequences as mentioned in the specification, or its complement), an isolated antibody, (or its antigen binding fragment) that binds to the above polypeptide, a hybridoma that produces the above monoclonal antibody, a pharmaceutical composition comprising the above antibody and a pharmaceutical excipient, a kit for detecting cancer cells (comprising the antibody cited above, methods for diagnosing cancer or for detecting the presence or absence of cancer cells in an individual, a method for inhibiting growth of cancer cells in an individual, a method for delivering a therapeutic agent to cancer cells in an individual, an electronic library comprising the above polynucleotide or polypeptide (or their fragments), methods of screening for anticancer activity or for a bioactive agent capable of modulating the activity of a CA protein (CAP), methods for detecting cancer associated with expression of a polypeptide in a test cell sample, a method for treating cancers and a method for inhibiting the expression of CA gene in a cell. The composition and methods are useful for detecting, diagnosing, preventing and treating cancers, especially lymphoma and leukemia. These may also be used in screening for agents that modulate cancer. The present sequence is a human CAP genomic sequence. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 277616 BP; 84510 A; 49827 C; 49846 G; 93433 T; 0 U; 0 Other;

Query Match 70.4%; Score 17.6; DB 13; Length 277616;  
Best Local Similarity 83.3%; Pred. No. 7.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTGCGCA 25  
DB 160286 AACTTAAGAGAAATATGTGCA 160309

RESULT 32  
ADA02624  
ID ADA02624 standard; DNA; 24888 BP.  
XX  
AC ADA02624;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Human NFKB1L1 carcinoma associated gene, SEQ ID NO:1142.  
XX  
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO2003057146-A2.  
XX  
PD 17-JUL-2003.  
XX  
PF 26-DEC-2002; 2002WO-US041414.  
XX  
PR 26-DEC-2001; 2001US-00035832.  
XX  
PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW;  
XX  
WPI; 2003-587068/55.  
XX  
DR New recombinant nucleic acid encoding carcinoma associated protein,  
PT useful for preparing compositions for treating carcinomas.  
XX  
PS Claim 1; SEQ ID NO 1142; 245pp; English.  
XX  
CC The invention relates to recombinant carcinoma associated (CA) nucleic  
CC acid sequences from mouse and human (ADA01482-ADA03094), and to  
CC recombinant carcinoma associated proteins (CAP) encoded by them. The  
CC invention also encompasses expression vectors and host cells comprising a  
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically  
CC binds to the protein, and a biochip comprising CA nucleic acid or  
CC fragments thereof. The sequences of the invention were identified using  
CC oncogenic retroviruses, which insert into the genome of the host organism  
CC at random. Many of these do not carry transduced host oncogenes or  
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
CC direct consequence of the effects of proviral integration into host  
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
CC leukaemia) or a propensity to carcinoma by determination of the sequence  
CC of a CA gene, or by determination of CA gene expression in particular  
CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
CC therapeutic agents and in screening and evaluating drug candidates. The  
CC present sequence represents a specifically claimed human CA nucleic acid  
CC sequence of the invention. Note: The complete sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 24888 BP; 6552 A; 6182 C; 5872 G; 6282 T; 0 U; 0 Other;

Query Match 69.6%; Score 17.4; DB 9; Length 24888;  
Best Local Similarity 94.7%; Pred. No. 7.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTGCG 20  
DB 1064 AAATTAAGAGACTATTGCG 1082

RESULT 34  
ADE95872  
ID ADE95872 standard; DNA; 24888 BP.  
XX  
AC ADE95872;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Human NFKB1L1 gene genomic DNA sequence.  
XX  
KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;

QY 2 AAATTAAGAGACTATTGCG 20  
DB 1064 AAATTAAGAGACTATTGCG 1082

RESULT 33  
ADB72362  
ID ADB72362 standard; DNA; 24888 BP.  
XX  
AC ADB72362;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Human NFKB1L1 gene.  
XX  
KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.  
XX  
OS Homo sapiens.  
XX  
PN WO2003008583-A2.  
XX  
PD 30-JAN-2003.  
XX  
PF 26-DEC-2001; 2001WO-US051291.  
XX  
PR 02-MAR-2001; 2001US-00798586.  
XX  
PR 23-OCT-2001; 2001US-00004113.  
XX  
PR 08-NOV-2001; 2001US-00052482.  
XX  
PR 30-NOV-2001; 2001US-00997722.  
XX  
PR 20-DEC-2001; 2001US-00034650.  
XX  
PA (SAGR-) SAGRES DISCOVERY.  
XX  
PI Morris DW, Engelhard EK;  
XX  
WPI; 2003-239337/23.  
XX  
DR New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
PT cancer, neoplasm, adenocarcinoma, or sarcomas.  
XX  
PS Claim 1; SEQ ID NO 190; 2304pp; English.  
XX  
CC The invention relates to a novel recombinant nucleic acid comprising a  
CC nucleotide sequence selected from any of the 660 sequences fully defined  
CC in the specification. A polynucleotide of the invention has cytostatic  
CC activity, and may have a use in gene therapy, or in a vaccine. The  
CC recombinant nucleic acids and polypeptides are useful for treating  
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
CC sarcomas. The present sequence represents a human gene of the invention.  
XX  
SQ Sequence 24888 BP; 6552 A; 6182 C; 5872 G; 6282 T; 0 U; 0 Other;

Query Match 69.6%; Score 17.4; DB 10; Length 24888;  
Best Local Similarity 94.7%; Pred. No. 7.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTGCG 20  
DB 1064 AAATTAAGAGACTATTGCG 1082

RESULT 34  
ADE95872  
ID ADE95872 standard; DNA; 24888 BP.  
XX  
AC ADE95872;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Human NFKB1L1 gene genomic DNA sequence.  
XX  
KW cancer diagnosis; cancer treatment; carcinoma; cytostatic; gene therapy;

KW lymphoma; breast cancer; prostate cancer; leukaemia; ds; human; NFKBIL1.  
XX Homo sapiens.  
XX WO2003039484-A2.  
XX PD 15-MAY-2003.  
XX 08-NOV-2002; 2002WO-US036071.  
XX 08-NOV-2001; 2001US-00052482.  
XX (SAGR-) SAGRES DISCOVERY.  
XX Morris DW, Engelhard EK;  
XX WPI; 2003-441462/41.  
XX New carcinoma associated nucleic acids and proteins, useful for screening  
PT drug candidates, or for diagnosing and treating carcinomas, e.g.  
PT lymphoma, breast cancer, prostate cancer or leukemia.  
XX Claim 1; SEQ ID NO 130; 793pp; English.  
PS  
XX This invention relates to novel recombinant nucleic acids for use in  
XX diagnosis and treatment of cancer, especially carcinomas, as well as the  
XX use of compositions in screening methods. The compositions of the  
XX invention may have cytostatic activity whilst the disclosed sequences may  
XX be useful for gene therapy. The carcinoma associated nucleic acids and  
XX proteins are useful for diagnosing and treating carcinomas, for example  
XX lymphoma, breast cancer, prostate cancer or leukaemia, or for screening  
XX drug candidates or bioactive agents capable of binding to, or modulating  
XX the activity of, a carcinoma associated protein. The present sequence is  
XX the genomic DNA sequence of the human NFKBIL1 gene which is a carcinoma  
XX associated gene of the invention.  
XX Sequence 24888 BP; 6552 A; 6182 C; 5872 G; 6282 T; 0 U; 0 Other;  
SQ  
Query Match 69.6%; Score 17.4; DB 10; Length 24888;  
Best Local Similarity 94.7%; Pred. No. 7.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAATTAAGAGACTATTTCG 20  
Db 1064 AAATTAAGAGACTATTGG 1082  
RESULT 35  
ADC56843/C  
ID ADC56843 standard; DNA; 50000 BP.  
XX  
AC ADC56843;  
XX  
XX 18-DEC-2003 (first entry)  
DT  
XX Human IKBL gene containing SNPs used to diagnose rheumatoid arthritis.  
DE  
XX rheumatoid arthritis; single nucleotide polymorphism; SNP; human; IKBL;  
KW ds.  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FH replace(16433,t)  
FT /\*tag= a  
FT /standard\_name= "Single nucleotide polymorphism"  
FT replace(19049,g)  
FT /\*tag= b  
FT /standard\_name= "Single nucleotide polymorphism"  
XX  
XX JP2002355069-A.  
PN  
XX 10-DEC-2002.  
PD

XX 10-DEC-2001; 2001JP-00376275.  
XX  
XX 12-DEC-2000; 2000JP-00378107.  
XX  
XX (INOK/) INOKO H.  
PA  
XX WPI; 2003-472592/45.  
XX  
XX Diagnosing rheumatoid arthritis by detecting a new genetic polymorphism.  
XX Claim 1; SEQ ID NO 1; 44pp; Japanese.  
XX  
XX This invention relates to a novel method of diagnosing rheumatoid  
XX arthritis by detecting single nucleotide polymorphisms (SNPs) at  
XX positions 16433, 18366, 19048. Specifically, it refers to SNPs A16433T,  
XX T18366C and C19048G. Furthermore, the present invention describes PCR  
XX primers that bind to the 3 base positions cited above and as such can be  
XX used to amplify these regions accordingly. The method is used for  
XX diagnosing rheumatoid arthritis. This polymucleotide sequence is the  
XX human IKBL gene containing SNPs of the invention.  
XX Sequence 50000 BP; 12405 A; 11455 C; 12071 G; 14069 T; 0 U; 0 Other;  
SQ  
Query Match 69.6%; Score 17.4; DB 10; Length 50000;  
Best Local Similarity 94.7%; Pred. No. 7.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAATTAAGAGACTATTTCG 20  
Db 25310 AAATTAAGAGACTATTGG 25292  
RESULT 36  
ADA03044/C  
ID ADA03044 standard; DNA; 53106 BP.  
XX  
XX ADA03044;  
XX  
XX 06-NOV-2003 (first entry)  
DT  
XX Human BAT1 carcinoma associated gene, SEQ ID NO:1562.  
DE  
XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;  
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;  
KW Gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO2003057146-A2.  
PN  
XX 17-JUL-2003.  
PD  
XX  
XX 26-DEC-2002; 2002WO-US041414.  
PF  
XX  
XX 26-DEC-2001; 2001US-00035832.  
PR  
XX (SAGR-) SAGRES DISCOVERY.  
PA  
XX Morris DW;  
XX  
XX WPI; 2003-587068/55.  
XX  
XX New recombinant nucleic acid encoding carcinoma associated protein,  
XX useful for preparing compositions for treating carcinomas.  
XX  
XX Claim 1; SEQ ID NO 1562; 245pp; English.  
PS  
XX The invention relates to recombinant carcinoma associated (CA) nucleic  
XX acid sequences from mouse and human (ADA01482-ADA03094), and to  
XX recombinant carcinoma associated proteins (CAP) encoded by them. The  
XX invention also encompasses expression vectors and host cells comprising a  
XX CA nucleic acid, a polypeptide (especially an antibody) that specifically  
XX

CC binds to the protein, and a biochip comprising CA nucleic acid or  
 CC fragments thereof. The sequences of the invention were identified using  
 CC oncogenic retroviruses, which insert into the genome of the host organism  
 CC at random. Many of these do not carry transduced host oncogenes or  
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a  
 CC direct consequence of the effects of proviral integration into host  
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose  
 CC carcinoma (especially breast cancer, prostate cancer, lymphoma or  
 CC leukemia) or a propensity to carcinoma by determination of the sequence  
 CC of a CA gene, or by determination of CA gene expression in particular  
 CC tissues. CA nucleic acids, proteins and antibodies are also useful as  
 CC therapeutic agents and in screening and evaluating drug candidates. The  
 CC present sequence represents a specifically claimed human CA nucleic acid  
 CC sequence of the invention. Note: The complete sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 53106 BP; 13295 A; 11886 C; 13365 G; 14560 T; 0 U; 0 Other;

Query Match 69.6%; Score 17.4; DB 9; Length 53106;  
 Best Local Similarity 94.7%; Pred. No. 7.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCG 20  
 |||||  
 Db 47351 AAATTAAGAGACTATTTCG 47333

RESULT 37  
 ADA66328/C  
 ID ADA66328 standard; DNA; 53106 BP.  
 XX  
 AC ADA66328;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Human BAT1 gene genomic DNA sequence.

XX carcinoma-associated gene; CA gene; Rorc gene; MCG15938 gene; BAT1 gene;  
 KW Iqgap1 gene; IQGAP1 gene; Zpf29 gene; hCG27579 gene; Kcnj9 gene;  
 KW KCNJ9 gene; Ppp3cc gene; PPP3CC gene; MCG9110 gene; hCG27579 gene;  
 KW cancer cell; lymphatic cell; breast cell; prostate cell; epithelial cell;  
 KW carcinoma-associated protein; CAP; cytostatic; gene therapy; anticancer;  
 KW vaccine; carcinoma; lymphoma carcinoma; lymphatic cancer; breast cancer;  
 KW prostate cancer; DNA vaccine; animal model; human; ds; BAT1.

XX Homo sapiens.  
 XX  
 XX WO20003053224-A2.  
 XX  
 XX 03-JUL-2003.  
 XX  
 XX 20-DEC-2002; 2002WO-US041776.  
 XX  
 XX 20-DEC-2001; 2001US-00034650.  
 XX  
 XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;  
 XX  
 XX WPI; 2003-569168/53.

XX Novel recombinant carcinoma-associated nucleic acid, useful for  
 PT evaluating the effect of a candidate carcinoma drug, and for diagnosing  
 PT carcinoma.

XX Claim 1; Page 67-75; 229pp; English.

XX This invention relates to a novel recombinant carcinoma-associated (CA)  
 CC nucleic acid comprising a fully defined genomic, mRNA or coding sequences  
 CC of mouse Rorc gene or human RORC gene, mouse MCG15938 or human gene BAT1,  
 CC mouse Iqgap1 gene or human IQGAP1 gene, mouse Zpf29 gene or human

CC hCG27579 gene, mouse Kcnj9 gene or human KCNJ9 gene, mouse Ppp3cc gene or  
 CC human PPP3CC gene, mouse MCG9110 gene or human hCG27579 gene, as given in  
 CC the specification. CA genes are genes which are preferably expressed in  
 CC cancer cells, preferably lymphatic, breast, prostate or epithelial cells.  
 CC A compound which modifies the expression of the CA genes or bind to  
 CC carcinoma-associated proteins (CAP) may have cytostatic activity and the  
 CC sequences of the invention may enable the use of gene therapy or a  
 CC development of an anticancer vaccine. Therefore the invention may be  
 CC useful for diagnosis and treatment of carcinomas, especially lymphoma  
 CC carcinoma, breast cancer and prostate cancer. The CA genes may also be  
 CC useful as DNA vaccines and for generating animal models of carcinomas.  
 CC The present sequence is that of the human BAT1 gene genomic DNA sequence  
 CC of the invention.

XX Sequence 53106 BP; 13295 A; 11886 C; 13365 G; 14560 T; 0 U; 0 Other;

Query Match 69.6%; Score 17.4; DB 9; Length 53106;  
 Best Local Similarity 94.7%; Pred. No. 7.9e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCG 20  
 |||||  
 Db 47351 AAATTAAGAGACTATTTCG 47333

RESULT 38  
 ADB72782/C  
 ID ADB72782 standard; DNA; 53106 BP.

XX  
 AC ADB72782;  
 XX

DT 04-DEC-2003 (first entry)

XX Human BAT1 gene.

XX human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;  
 KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Homo sapiens.

XX WO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001WO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,  
 PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 610; 2304pp; English.

XX The invention relates to a novel recombinant nucleic acid comprising a  
 CC nucleotide sequence selected from any of the 660 sequences fully defined  
 CC in the specification. A polynucleotide of the invention has cytostatic  
 CC activity, and may have a use in gene therapy, or in a vaccine. The  
 CC recombinant nucleic acids and polypeptides are useful for treating  
 CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and  
 CC sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 53106 BP; 13295 A; 11886 C; 13365 G; 14560 T; 0 U; 0 Other;

```

Query Match          69.6%; Score 17.4; DB 10; Length 53106;
Best Local Similarity 94.7%; Pred. No. 7.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTGC 20
   |||||
DB 47351 AAATTAAGAGACTATTGG 47333

RESULT 39
ADL27122/c
ID ADL27122 standard; DNA; 53106 BP.
XX
AC ADL27122;
XX
DT 20-MAY-2004 (first entry)
XX
DE Human genomic sequence for BAT1.
XX
KW Human; ds; gene; CA gene; carcinoma associated gene; cytostatic; cancer;
KW carcinoma; lymphoma.
XX
OS Homo sapiens.
XX
PN US2003216558-A1.
XX
PD 20-NOV-2003.
XX
PF 20-DEC-2001; 2001US-00034650.
XX
PR 22-DEC-2000; 2000US-00747377.
XX
PR 02-MAR-2001; 2001US-00798586.
XX
PR
XX
PA (MORR/) MORRIS D W.
PA (ENGE/) ENGELHARD E K.
XX
XX
PI Morris DW, Engelhard EK;
XX
XX
DR WPI; 2003-902052/82.
XX
XX
PT New carcinoma associated gene, useful for preparing a composition for
PT diagnosing or treating carcinoma.
XX
PS Claim 1; SEQ ID NO 10; 342pp; English.
XX
XX
CC The invention relates to a new recombinant nucleic acid (from a CA,
CC carcinoma associated, gene) appearing as ADL27113-ADL27172. Also included
CC are a host cell comprising the recombinant nucleic acid or expression
CC vector, an expression vector comprising the recombinant nucleic acid, a
CC recombinant protein (a carcinoma associated protein) comprising the
CC sequence encoded by the nucleic acid, a method for screening drug
CC candidates, a method for screening for a bioactive agent capable of
CC binding to (or modulating the activity of) a carcinoma associated
CC protein, a method for evaluating the effect of a candidate carcinoma
CC drug, a method of diagnosing carcinoma, a method for inhibiting the
CC activity of a carcinoma associated protein, a method of treating
CC carcinomas, a method of neutralising the effect of a carcinoma associated
CC protein, a biochip comprising one or more nucleic acid segments of the
CC nucleic acid, a method of diagnosing carcinoma or propensity to carcinoma
CC and a method of determining carcinoma associated gene copy number. The
CC nucleic acid is useful for preparing a composition for diagnosing or
CC treating carcinoma especially lymphomas. The present sequence is the
CC genomic sequence from a human carcinoma associated gene.
XX
SQ Sequence 53106 BP; 13296 A; 11983 C; 13268 G; 14559 T; 0 U; 0 Other;

Query Match          69.6%; Score 17.4; DB 11; Length 53106;
Best Local Similarity 94.7%; Pred. No. 7.9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTGC 20
   |||||
DB 47351 AAATTAAGAGACTATTGG 47333

```

```

RESULT 40
ABK84756/c
ID ABK84756 standard; cDNA; 81800 BP.
XX
AC ABK84756;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #1327.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
PN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PR
XX
PA (GENE-) GENE LOGIC INC.
XX
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
XX
DR WPI; 2002-435328/46.
XX
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 1327; 114pp; English.
XX
XX
CC The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease; also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above

```



CC conditions. The present sequence represents a gene differentially  
CC expressed in granulocytes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 81800 BP; 19886 A; 19955 C; 20623 G; 21336 T; 0 U; 0 Other;  
Query Match 69.6%; Score 17.4; DB 6; Length 81800;  
Best Local Similarity 94.7%; Pred. No. 8.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTTCG 20  
DB 80932 AAATTAAAGAGACTATTTCG 80914

RESULT 41  
AAH41224  
ID AAH41224 standard; DNA; 349980 BP.  
XX  
AC AAH41224;  
XX  
DT 29-OCT-2001 (first entry)  
XX  
XX Pyrococcus abyssi genomic fragment #3.  
DE  
XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.  
KW  
XX Pyrococcus abyssi.  
OS  
XX

FH Key Location/Qualifiers  
FT misc\_feature 1..49980  
FT /\*tag= a  
FT /note= "This sequence overlaps with the 3' end of  
FT AAH41223"  
FT misc\_feature 300001..349980  
FT /\*tag= b  
FT /note= "This sequence overlaps with the 5' end of  
FT AAH41225"  
XX  
XX  
XX FR2792651-A1.  
XX  
XX 27-OCT-2000.  
XX  
XX 21-APR-1999; 99FR-00005034.  
XX  
XX 21-APR-1999; 99FR-00005034.  
XX  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX (IFRS-) IFREMER INST FR RECH EXPL MER.  
XX  
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;  
XX Querellou J, Weissenbach J, Saurin W, Heilig R;  
XX WPI; 2001-126236/14.  
XX  
XX New nucleotide sequences isolated from Pyrococcus abyssi encode proteins  
XX useful in industry.  
XX  
XX Claim 1; Page 347-443; 1657pp; French.

XX The present invention relates to the genomic sequence of Pyrococcus  
CC abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a  
CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal  
CC vents. The present sequence is a fragment of the genomic sequence of P.  
CC abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41223  
CC and the 3' end of this sequence overlaps with the 5' end of AAH41225. The  
CC proteins of the present invention have various potential industrial uses,  
CC since the proteins are stable at very high temperatures, some up to 110  
CC degrees centigrade. Note: This patent is in the same patent family as  
CC WO200065062, which contains additional sequences as shown in AAB99132-  
CC AAB99143, AAH75903-AAH75920 and AAG66436

XX SQ Sequence 349980 BP; 94090 A; 78692 C; 78319 G; 98879 T; 0 U; 0 Other;  
Query Match 69.6%; Score 17.4; DB 5; Length 349980;  
Best Local Similarity 94.7%; Pred. No. 9.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTTCG 20  
DB 72176 AAATTAAAGAGACTATTTCG 72194

RESULT 42  
AAC28412  
ID AAC28412 standard; cDNA; 490 BP.  
XX  
AC AAC28412;  
XX  
XX 06-OCT-2000 (first entry)  
XX  
XX Human secreted protein 5' EST, SEQ ID NO: 32487.  
DE  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-00200610.  
XX  
XX 26-FEB-1999; 99US-0122487P.  
XX (GEST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX WPI; 2000-500381/45.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.  
XX Claim 1; SEQ ID NO 32487; 71pp + Sequence Listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
CC identified within the present sequence. The 5' ESTs were prepared from  
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
CC derived from the 5' ends of mRNAs and even in those cases where longer  
CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
CC They are used to obtain upstream regulatory sequences and to design  
CC expression and secretion vectors

XX SQ Sequence 490 BP; 136 A; 93 C; 112 G; 149 T; 0 U; 0 Other;  
Query Match 68.8%; Score 17.2; DB 3; Length 490;  
Best Local Similarity 86.4%; Pred. No. 6.3e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTCAA 25  
DB 377 ATTAAGAGACTATTTCGTCAA 398





```
PN WO2005052145-A2.
XX
XX
XX 09-JUN-2005.
XX
XX
XX 07-OCT-2004; 2004WO-US033214.
XX
XX 07-OCT-2003; 2003US-0509175P.
XX
XX (UYFL ) UNIV FLORIDA.
XX
XX Burne RA, Yi-Ywan MC;
XX
XX WPI; 2005-417990/42.
XX
XX Novel recombinant bacterial cell comprising isolated nucleic acid
PT construct, and expressing one or more alkalizing enzyme, useful for
PT treating dental caries.
XX
XX Claim 25; SEQ ID NO 2; 135pp; English.
XX
XX The invention relates to genetically engineered bacteria that promote
CC alkalization of dental biofilms. The bacteria comprise nucleic acid
CC constructs having components necessary to express functional alkali-
CC producing enzymes, such as urease, arginine deiminase and agmatine
CC deiminase. Following introduction into the oral cavity and colonization
CC of the plaque, the bacteria produce alkalis such as ammonia from urea,
CC arginine or agmatine, thereby raising the pH and reducing the incidence
CC of caries. The bacterial strain is a dental plaque-colonizing bacterium
CC such as Streptococcus mutans, Streptococcus sanguinis, Streptococcus
CC gordonii, Streptococcus parasanguis, Streptococcus vestibularis,
CC Streptococcus oralis and Streptococcus mitis. In preferred embodiments,
CC the nucleic acid construct comprises a gene cluster encoding a urease and
CC nickel transporter such as ureIABCFGDMQO, an arginine deiminase system
CC such as arcABCDTR, or an agmatine deiminase system such as aquBAC and a
CC transcriptional regulator located upstream of the agu gene cluster, or a
CC combination of these. A vector comprising the nucleic acid construct is
CC designed to target a mtl gene. The vector is inserted into the mtl gene
CC by double-crossover recombination, and the host cell is selected by its
CC inability to grow on mannitol. A composition comprising the recombinant
CC bacteria may also comprise a carrier such as a chewing gum, toothpaste or
CC lozenge. The present sequence is that of the arginine deiminase gene
CC cluster from Streptococcus gordonii strain DLI, including an anaerobic
CC regulator, flp. The gene cluster comprises open reading frames designated
CC arcA (arginine deiminase), arcB (ornithine carbamyltransferase), arcC
CC (carbamate kinase), arcD (arginine-ornithine antitransporter), arcE
CC (dipeptidase) and arcX (regulator).
XX
XX Sequence 9398 BP; 2883 A; 1709 C; 1983 G; 2823 T; 0 U; 0 Other;
SQ
Query Match 68.8%; Score 17.2; DB 14; Length 9398;
Best Local Similarity 86.4%; Pred. No. 8.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CAAATTAAAGAGACTATTCTGTG 22
Db 1448 CAAACTAAAGTGGCTATTCTGTG 1469
RESULT 46
ABA88966/c
ID ABA88966 standard; DNA; 12264 BP.
XX
XX ABA88966;
AC
XX
XX 11-FEB-2002 (first entry)
DT
XX
XX Escherichia coli polynucleotide SEQ ID NO 507.
DE
XX
XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicaemia;
KW pyelonephritis; antibiotic resistance; ds.
XX
```

```
OS Escherichia coli.
XX
XX WO200166572-A2.
XX
XX 13-SEP-2001.
XX
XX 12-MAR-2001; 2001WO-EP003445.
XX
XX 10-MAR-2000; 2000FR-00003145.
XX 02-FEB-2001; 2001FR-00001449.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX WPI; 2001-550253/61.
XX
XX A library of DNA fragments of Escherichia coli strains for the phylogenic
PT determination of a given strain comprises polynucleotides of nature B2/D+
PT A-.
XX
XX Example 6; Fig 6; 646pp; English.
XX
XX The invention relates to a library of DNA fragments of Escherichia coli
XX strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) and
XX encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
XX B2/D+A-. The polynucleotides have potential antiinflammatory,
XX antibacterial and immunosuppressive activity as part of pharmaceutical
XX compositions used to treat, palliate or prevent extra-intestinal E. coli
XX infections. The polypeptides are useful for determining the phylogenic
XX group of a given E. coli strain. These polypeptides can detect and treat
XX an undesired development of E. coli, particularly an extra-intestinal
XX infection that include systemic and non-diarrhoeal infections such as
XX septicaemia, pyelonephritis and meningitis this is particularly
XX advantageous as bacterial resistance is increasing with the more frequent
XX use of broad spectrum antibiotics
XX
XX Sequence 12264 BP; 3331 A; 2836 C; 2579 G; 3518 T; 0 U; 0 Other;
SQ
Query Match 68.8%; Score 17.2; DB 4; Length 12264;
Best Local Similarity 86.4%; Pred. No. 8.5e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAATTAAAGAGACTATTCTGTG 23
Db 7425 AAATTAAATGATACTATTCTGTG 7404
RESULT 47
ABS78844
ID ABS78844 standard; DNA; 23654 BP.
XX
XX ABS78844;
AC
XX
XX 17-DEC-2002 (first entry)
DT
XX
XX E. coli CFT073 genomic sequence #11.
DE
XX
XX Pathogenic; Escherichia coli CFT073 infection; livestock; pyelonephritis;
KW urinary tract infection; open reading frame; ORF; uropathogenic;
KW antibacterial; atrophic; nephrotropic; gene; ds.
XX
XX Escherichia coli.
OS
XX WO200259320-A2.
XX
XX 01-AUG-2002.
XX
XX 19-OCT-2001; 2001WO-US046833.
XX
XX 19-OCT-2000; 2000US-0242412P.
XX
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
```

XX Blattner FR, Welch RA, Burland VD;  
PI WPI; 2002-691532/74.  
XX New DNA sequences of the pathogenic *Escherichia coli* CFT073 strain,  
PT useful for preventing or treating *E. coli* CFT073 infection in humans or  
PT livestock.  
XX Claim 1; Page 17-30; 765pp; English.  
XX The present invention relates to polynucleotide sequences from the genome  
CC of the pathogenic *Escherichia coli* strain CFT073. Almost all the  
CC sequences present in *E. coli* CFT073 are absent in the previously  
CC sequenced laboratory strain K-12. The polynucleotide sequences of the  
CC invention are useful for preventing, diagnosing or treating *E. coli*  
CC CFT073 infection in humans or livestock. The polynucleotide sequences are  
CC useful for preventing urinary tract infections and pyelonephritis.  
CC Likewise, the polypeptides encoded by the different open reading frames  
CC (ORF1-5) are useful for generating a vaccine against uropathogenic *E.*  
CC *coli* strains. ABS78834-ABS79085 represent genomic sequences from *E. coli*  
CC strain CFT073  
XX  
SQ Sequence 23654 BP; 6184 A; 5664 C; 5995 G; 5806 T; 0 U; 5 Other;  
Query Match 68.8%; Score 17.2; DB 6; Length 23654;  
Best Local Similarity 86.4%; Pred. No. 9e+02; 3; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAATTAAAGAGACTATTCTGTC 23  
Db 4839 AAATTAAAGAGACTATTCTGTC 4860  
RESULT 48  
ID ADH80411 standard; DNA; 23654 BP.  
XX  
AC ADH80411;  
XX  
DT 22-APR-2004 (first entry)  
DE *Escherichia coli* CFT073 genome contig #11.  
XX  
XX ds; gene; *Escherichia coli*; CFT073; *Escherichia coli* CFT073 infection.  
XX  
XX *Escherichia coli*; CFT073.  
XX  
XX US2003165870-A1.  
XX  
XX 04-SEP-2003.  
XX  
XX 01-MAR-2002; 2002US-00085959.  
XX  
XX 01-MAR-2002; 2002US-00085959.  
XX  
XX (BLAT/) BLATTNER F R.  
XX (WELC/) WELCH R A.  
XX (BURL/) BURLAND V D.  
XX  
XX Blattner FR, Welch RA, Burland VD;  
XX  
XX WPI; 2003-863698/80.  
XX  
XX New nucleic acid of *Escherichia coli* CFT073, useful for preparing a  
PT composition for diagnosing, treating or preventing infection caused by  
PT *Escherichia coli* CFT073.  
XX  
XX Claim 1; SEQ ID NO 11; 4pp; English.  
XX  
XX The invention relates to an isolated *Escherichia coli* CFT073 nucleic acid  
CC molecule. The nucleic acid is useful for preparing a composition for  
CC diagnosing, treating or preventing infection caused by *Escherichia coli*

CC CFT073. The present sequence represents a contig of the *Escherichia coli*  
XX CFT073 genome that is not present in *Escherichia coli* K-12.  
SQ Sequence 23654 BP; 6184 A; 5664 C; 5995 G; 5806 T; 0 U; 5 Other;  
Query Match 68.8%; Score 17.2; DB 10; Length 23654;  
Best Local Similarity 86.4%; Pred. No. 9e+02; 3; Indels 0; Gaps 0;  
Matches 19; Conservative 0; Mismatches 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 AAATTAAAGAGACTATTCTGTC 23  
Db 4839 AAATTAAAGAGACTATTCTGTC 4860  
RESULT 49  
ID AAL35943 standard; DNA; 32132 BP.  
XX  
XX AAL35943;  
XX  
DT 08-JAN-2002 (first entry)  
XX  
DE Human musculoskeletal system related polynucleotide SEQ ID NO 2308.  
XX  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein;  
KW musculoskeletal system; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155367-A1.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001338.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 18-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226681P.  
XX 22-AUG-2000; 2000US-0226686P.  
XX 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.

PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Barash SC, Ruben SM;

XX WPI; 2001-451937/48.

PT Isolated polypeptide for treating, preventing and/ or prognosing  
PT disorders related to the musculoskeletal system including musculoskeletal  
PT cancers and also for testing and detection e.g. diagnosis.

XX Example 2; SEQ ID NO 2308; 781pp + Sequence Listing; English.

XX The invention relates to novel genes (AAL34669-AAL37666) and proteins  
CC (ABE03087-ABE04109) associated with the musculoskeletal system useful for  
CC preventing, treating or ameliorating medical conditions e.g. by protein  
CC or gene therapy. The genes are isolated from a range of human tissues  
CC disclosed in the specification. The nucleic acids, proteins, antibodies  
CC and (ant)agonists are useful in the diagnosis, treatment and prevention  
CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,  
CC lung, or urogenital; (b) immune disorders e.g. Addison's disease,  
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid  
CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as  
CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.  
CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,  
CC bacterial, fungal and parasitic infections. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 4; Length 32132;  
Best Local Similarity 86.4%; Pred. No. 9.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTTAAGAGACTATTCGTGCAA 25

Db 26595 ATTTAAGAGATAATTCATGCAA 26574

RESULT 50

AAL07060/c  
ID AAL07060 standard; DNA; 32132 BP.  
XX  
AC AAL07060;  
DT 21-NOV-2001 (first entry)  
XX  
XX Human reproductive system related antigen DNA SEQ ID NO: 9748.  
DE  
XX Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200155320-A2.  
XX  
XX 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US0001339.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0234844P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0241826P.  
PR 08-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.

PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
PA (HUNA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
DR  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
PT  
XX  
XX Disclosure; SEQ ID NO 9748; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention  
XX  
XX Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;  
SQ  
Query Match 68.8%; Score 17.2; DB 4; Length 32132;  
Best Local Similarity 86.4%; Pred. No. 9.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4 ATTAAGAGACTATTGTCGAA 25  
||||||| |||||  
Db 26595 ATTAAGAGACTATTGTCGAA 26574  
RESULT 51  
ABX58931/c  
ID ABX58931 standard; cDNA; 32132 BP.  
XX  
AC ABX58931;  
XX  
XX 26-FEB-2003 (first entry)  
XX  
XX cDNA encoding novel human musculoskeletal system antigen #1275.  
XX  
KW Gene; ss; musculoskeletal system antigen; cancer; metastasis;  
KW re-vascularisation; thrombosis; arteriosclerosis; mineral content;  
KW cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;  
KW post-operative tissue repair; limb regeneration; neuronal growth;  
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
KW AIDS-related complex; chondrocyte growth; bone regeneration;  
KW periodontal regeneration; tissue transport; bone graft; skin aging;  
KW keratinocyte growth; hair loss; melanocyte growth; cell proliferation;  
KW cell growth; organ transplant; cell differentiation; body height; weight;  
KW hair colour; eye colour; skin; percentage of adipose tissue;  
KW pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm;  
KW depression; tendency for violence; pain; reproductive capability;  
KW hormone level; endocrine level; appetite; libido; memory; stress;  
KW storage capability; fat content; lipid content; protein content;  
KW carbohydrate content; vitamin content; cofactor content;  
KW nutritional component.  
XX  
XX Homo sapiens.  
XX  
XX US2002147140-A1.  
PN  
XX  
XX 10-OCT-2002.  
PD  
XX  
XX 17-JAN-2001; 2001US-00764877.  
PF

XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180638P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 26-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 22-AUG-2000; 2000US-0225758P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX  
XX (ROSE/) ROSEN C A.  
PA (RUBE/) RUBEN S M.  
XX (BARA/) BARASH S C.  
XX  
XX Rosen CA, Ruben SM, Barash SC;  
PI  
XX WPI; 2003-128199/12.  
XX  
XX Isolated nucleic acid molecules encoding musculoskeletal system  
PT associated polypeptides, useful for detecting disorders, e.g. cancer.  
XX  
XX Disclosure; SEQ ID NO 2308; 321pp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule comprising a  
CC sequence encoding musculoskeletal system associated polypeptides useful  
CC for detecting disorders, e.g., cancer or cancer metastases, in animals or  
CC humans. The nucleic acid: stimulates re-vascularisation of ischaemic  
CC tissues associated with conditions such as thrombosis, arteriosclerosis,  
CC and other cardiovascular conditions; treats wounds due to injuries,  
CC burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
CC and limb regeneration; stimulates neuronal growth; can treat and prevent

CC neuronal damage occurring in certain disorders or neurodegenerative  
CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-  
CC related complex; stimulates chondrocyte growth, thus they can be used to  
CC enhance bone and periodontal regeneration and aid in tissue transports or  
CC bone grafts; prevents skin aging due to sunburn by stimulating  
CC keratinocyte growth; prevents hair loss, since FGF family members  
CC activate hair-forming cells and promotes melanocyte growth; stimulates  
CC growth and differentiation of hematopoietic cells and bone marrow cells  
CC when used in combination with other cytokines; maintains organs before  
CC transplantation or for supporting cell culture of primary tissues;  
CC induces tissue of mesodermal origin to differentiate in early embryos;  
CC increases or decreases the differentiation or proliferation of embryonic  
CC stem cells, besides, haematopoietic lineage; modulates mammalian  
CC characteristics, such as, body height, weight, hair colour, and shape (e.g.,  
CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal  
CC state or physical state by influencing biorhythms, circadian rhythms,  
CC depression, tendency for violence, tolerance for pain, reproductive  
CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or  
CC stress; increases or decreases storage capabilities, fat content, lipid,  
CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
CC components. This sequence encodes a novel human musculoskeletal system  
CC antigen. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from the US patent office at  
CC ftp.seqdata.uspto.gov/sequence.html?DocID=20020147140

XX Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 8; Length 32132;  
Best Local Similarity 86.4%; Pred. No. 9.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCGTGCAA 25  
||||||| |||||  
Db 26595 ATTAAGAGATAATTCATGCAA 26574

## RESULT 52

ADJ29681/c

ID ADJ29681 standard; DNA; 32132 BP.

AC ADJ29681;

XX 20-MAY-2004 (first entry)

DT Human musculoskeletal system-associated genomic DNA - SSQ ID 2308.

DE musculoskeletal system; cytostatic; osteopathic; cancer; osteoporosis;

KW gene therapy; vaccine; human; ds.

XX Homo sapiens.

XX US2004009488-A1.

XX 15-JAN-2004.

XX 13-SEP-2002; 2002US-00242515.

XX 31-JAN-2000; 2000US-0179065P.

XX 04-FEB-2000; 2000US-0180628P.

XX 24-FEB-2000; 2000US-0184684P.

XX 02-MAR-2000; 2000US-0186350P.

XX 16-MAR-2000; 2000US-0189874P.

XX 17-MAR-2000; 2000US-0190076P.

XX 18-APR-2000; 2000US-0198123P.

XX 19-MAY-2000; 2000US-0205515P.

XX 07-JUN-2000; 2000US-0209467P.

XX 28-JUN-2000; 2000US-0214886P.

XX 30-JUN-2000; 2000US-0215135P.

XX 07-JUL-2000; 2000US-0216647P.

XX 07-JUL-2000; 2000US-0216880P.

XX 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 25-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.

PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 06-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PR 17-JAN-2001; 2001US-00764877.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PA  
XX Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2004-090458/09.  
XX  
PT New nucleic acid molecule, useful for preparing a medicament for  
PT preventing, treating or ameliorating a medical condition e.g., cancer of  
PT musculoskeletal tissues or osteoporosis.  
XX  
PS Disclosure; SEQ ID NO 2308; 289pp; English.  
XX  
CC The invention relates to a novel isolated musculoskeletal system-  
CC associated nucleic acid molecule. The nucleic acid of the invention  
CC demonstrates cytostatic and osteopathic activities and may be useful for  
CC preparing a medicament for preventing, treating or ameliorating a medical  
CC condition such as cancer of the musculoskeletal tissues or osteoporosis,  
CC possibly via gene therapy or vaccine production. The current sequence is  
CC that of the human musculoskeletal system-associated genomic DNA of the  
CC invention. The current sequence is not shown within the specification per  
CC se but is available on the USPTO web-site  
CC http:seqdata.uspto.gov/sequence.html?DocID=20040009488.

XX  
SQ Sequence 32132 BP; 8574 A; 7006 C; 7175 G; 9377 T; 0 U; 0 Other;  
Query Match 68.8%; Score 17.2; DB 12; Length 32132;  
Best Local Similarity 86.4%; Pred. No. 9.2e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 4 ATTAAGAGAGACTATTTCGTGCAA 25  
|||||  
Db 26595 ATTAAGAGAGATAATTCATGCAA 26574  
|||||  
RESULT 53  
ACN43998\_6/c  
Continuation (7 of 7) of ACN43998 from base 600001 (Human genomic sequence HCG1778483.)  
WP Sequence split into 7 fragments LOCUS ACN43998 Accession ACN43998  
WP Fragment Name Begin End  
WP ACN43998\_0 1 110000  
WP ACN43998\_1 100001 210000  
WP ACN43998\_2 200001 310000  
WP ACN43998\_3 300001 410000  
WP ACN43998\_4 400001 510000  
WP ACN43998\_5 500001 610000  
WP ACN43998\_6 600001 653122  
Query Match 68.8%; Score 17.2; DB 11; Length 53122;  
Best Local Similarity 86.4%; Pred. No. 9.7e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGAGACTATTTCGTG 22  
|||||  
Db 7207 CAAATTAAGAGAGAAATTTGTG 7186  
|||||  
RESULT 54  
ABZ74619/c  
ID ABZ74619 standard; DNA; 58181 BP.  
XX AC ABZ74619;  
XX  
DT 12-MAY-2003 (first entry)  
XX  
DE Secreted protein gene 372 genomic fragment HWBCN36, SEQ ID NO:1766.  
XX Human; secreted protein; cancer; tumour; hyperproliferative disorder;  
XX autoimmune disorder; inflammation; angiogenic diseases; AIDS;  
XX acquired immunodeficiency syndrome; hepatitis; anaemia; wound healing;  
XX drug screening; chromosome identification; chromosome mapping;  
XX cytostatic; gene therapy; antinflammatory; immunomodulator; anti-HIV;  
XX antianaemic; vulnery; gene; ds.  
XX OS Homo sapiens.  
XX  
XX PN WO200277013-A2.  
XX  
XX PD 03-OCT-2002.  
XX  
XX PF 26-MAR-2002; 2002WO-US009370.  
XX  
XX PR 27-MAR-2001; 2001US-0278650P.  
XX PR 12-SEP-2001; 2001US-00950082.  
XX PR 12-SEP-2001; 2001US-00950083.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX PA  
XX Rosen CA, Ruben SM;  
XX WPI; 2003-040578/03.  
XX  
XX New human secreted proteins and nucleic acids, useful for detecting or  
XX treating cancer or other hyperproliferative disorders, autoimmune  
XX disorders, inflammatory disorders, HIV disease, hepatitis or anemia.  
XX



PS Disclosure; Page 2331-2345; 2474pp; English.

XX ABZ73281-ABZ73697 represent cDNAs corresponding to 391 human secreted protein genes, and ABP00947-ABP01363 represent the proteins they encode.

CC ABZ73698-ABZ74687 represent human secreted protein genomic fragments. The invention also encompasses antibodies specific for the secreted proteins, the use of the secreted proteins in drug screening and recombinant vectors and host cells comprising a nucleic acid of the invention. The secreted proteins are thought to be involved in biological activities associated with cellular signalling, cellular differentiation, cell migration, pro hormone activation and neurotransmitter activity. The secreted proteins, nucleic acids encoding them, antibodies or antibody fragments specific for the secreted proteins, and modulators of protein activity are useful for diagnosing or treating cancers or other hyperproliferative disorders. Additionally, the secreted proteins and their nucleic acids may also be used in the treatment of autoimmune disorders, inflammatory disorders, diseases involving angiogenesis, AIDS (acquired immunodeficiency syndrome), hepatitis, anaemia, and to promote wound healing. Nucleic acids of the invention may be used for chromosome identification, chromosome mapping, in gene therapy, for identifying individuals from minute biological samples, as hybridisation probes, and as molecular weight markers. The present sequence represents a human secreted protein genomic fragment referred to in the disclosure of the invention

XX SQ Sequence 58181 BP; 15503 A; 12755 C; 13504 G; 16419 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 8; Length 58181;  
Best Local Similarity 86.4%; Pred. No. 9.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTGCAA 25  
|||||  
DZ 26595 ATTAAGAGATAATTTCATGCAA 26574

RESULT 55  
ADC21010/c  
ID ADC21010 standard; DNA; 58181 BP.  
XX AC ADC21010;  
XX DT 18-DEC-2003 (first entry)  
XX DE Human secreted protein-related DNA sequence #428.  
XX KW gene therapy; human; secreted protein; haemopoietic disorder;  
KW haematological disorder; anaemia; haemophilia; inflammatory disorder;  
KW inflammatory bowel disease; Crohn's disease; neoplastic disease; cancer;  
KW leukaemia; wound healing; epithelial cell proliferation disorder;  
KW immune disorder; autoimmune disorder; asthma; asthma;  
KW cardiovascular disorder; atherosclerosis; myocarditis;  
KW infectious disease; HIV; AIDS; endocrine disorder; diabetes;  
KW gastrointestinal disorder; duodenal ulcer; gastroenteritis; gene; ds.  
XX OS Homo sapiens.  
XX PN WO200292787-A2.  
XX PD 21-NOV-2002.  
XX PF 26-MAR-2002; 2002WO-US009257.  
XX PR 27-MAR-2001; 2001US-0278650P.  
XX PR 12-SEP-2001; 2001US-00950082.  
XX PR 12-SEP-2001; 2001US-00950083.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM;  
XX WPI; 2003-129287/12.

PT New human secreted proteins and nucleic acid molecules, useful for preparing a diagnostic or pharmaceutical composition for diagnosing, preventing or treating hematopoietic or hematologic disorders, e.g. anemia or hemophilia.

XX Disclosure; SEQ ID NO 964; 1512pp; English.

XX The invention comprises the amino acid and coding sequences of human secreted proteins. The DNA and protein sequences of the invention are useful for detecting, preventing, diagnosing, prognosticating, treating or ameliorating; hematopoietic or hematological disorders (e.g. anaemia and haemophilia); inflammatory disorders (e.g. inflammatory bowel disease and Crohn's disease); neoplastic disease (e.g. cancer and leukaemia); wound healing and disorders of epithelial cell proliferation; immune disorders (e.g. autoimmune disorders and asthmatic disorders); cardiovascular disorders (e.g. atherosclerosis and myocarditis); infectious diseases (e.g. HIV/AIDS); endocrine disorders (e.g. diabetes); and gastrointestinal disorders (e.g. duodenal ulcers and gastroenteritis). The present DNA sequence was used in the exemplification of the invention.

XX SQ Sequence 58181 BP; 15503 A; 12755 C; 13504 G; 16419 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 10; Length 58181;  
Best Local Similarity 86.4%; Pred. No. 9.8e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTGCAA 25  
|||||  
DZ 26595 ATTAAGAGATAATTTCATGCAA 26574

RESULT 56  
ABZ68140/c  
ID ABZ68140 standard; DNA; 58181 BP.  
XX AC ABZ68140;  
XX DT 26-MAR-2003 (first entry)  
XX DE Human secreted protein encoding genomic DNA SEQ ID NO 1663.  
XX KW Human; secreted protein; neurotropic; neuroprotective; cytostatic; virucide; dermatological; immunosuppressive; antiinflammatory; anti-HIV; vulnery; antibacterial; antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer; antirheumatic; hepatotropic; carebroprotective; antiinflammatory; antiallergic; antidiabetic; antileukemic; anticonvulsant; antifungal; antiparasitic; cardiac; immune disorder; infection; vaccine; cardiovascular disorder; neurological disease; nephrotropic; gene therapy; gene; ds.  
XX OS Homo sapiens.  
XX PN WO200277186-A2.  
XX PD 03-OCT-2002.  
XX PF 26-MAR-2002; 2002WO-US009188.  
XX PR 27-MAR-2001; 2001US-0278650P.  
XX PR 12-SEP-2001; 2001US-00950082.  
XX PR 12-SEP-2001; 2001US-00950083.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Ruben SM;  
XX WPI; 2003-040583/03.  
XX New human secreted proteins encoded by genes contained in cDNA clones (e.g. HGCAC19), useful for preventing, treating or diagnosing e.g. AIDS, multiple sclerosis, herpes virus, leukemia, tick-borne encephalitis or West Nile fever.

XX PS Disclosure; Page 2272-2286; 2423pp; English.

XX CC The invention relates to novel human genes (ABZ66891-ABZ68209) and the

XX CC encoded secreted proteins (ABP99470-ABP99872) useful for preventing,

XX CC treating or ameliorating medical conditions e.g. by protein or gene

XX CC therapy. The genes are isolated from a range of human tissues disclosed

XX CC in the specification. The nucleic acids, proteins, antibodies and

XX CC (ant)agonists are useful in the diagnosis, treatment and prevention of:

XX CC (a) cancer, e.g. breast and ovarian cancer and other cancers of the

XX CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,

XX CC lung or urogenital; (b) immune disorders e.g. Addison's disease,

XX CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,

XX CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid

XX CC arthritis and ulcerative colitis; (c) cardiovascular disorders such as

XX CC myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g.

XX CC cerebral anoxia and epilepsy; and (f) infectious diseases such as viral,

XX CC bacterial, fungal and parasitic infections

XX SQ Sequence 58181 BP; 15503 A; 12755 C; 13504 G; 16419 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 10; Length 58181;

Best Local Similarity 86.4%; Pred. No. 9.8e+02;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ATTAAGAGACTATTCGTGCAA 25

Db 26595 ATTAAGAGACTATTCGTGCAA 26574

RESULT 57

ADZ12890/c

ID ADZ12890 standard; DNA; 101636 BP.

AC ADZ12890;

XX 16-JUN-2005 (first entry)

DE Murine cancer-associated genomic DNA #37.

KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;

XX cystostatic; gene; ds.

OS Mus sp.

XX WO2005031001-A2.

PN 07-APR-2005.

XX 23-SEP-2004; 2004WO-US031617.

XX 23-SEP-2003; 2003US-00669920.

XX (CHIR ) CHIRON CORP.

XX Morris DW, Malandro MS;

XX WPI; 2005-273395/28.

XX Nucleic acid array useful for detecting cancer associated nucleic acid,

XX comprises two or more nucleic acid probes.

XX Disclosure; SEQ ID NO 410; 198pp; English.

XX The invention relates to a nucleic acid array for detecting a cancer

XX associated (CA) nucleic acid, comprising two or more nucleic acid probes.

XX The invention also relates to a peptide array comprising two or more

XX isolated polypeptides encoded by a CA nucleic acid sequence, a compound

XX that binds to a polypeptide, an isolated antibody or its fragment which

XX binds to a polypeptide, which is prepared by immunizing a host animal

XX with a composition comprising the polypeptide or its antigen binding

XX fragment and collecting cells from the host expressing antibodies against

XX the antigen or its antigen binding fragment, a composition comprising the

CC antibody and a carrier, a method of screening for anticancer activity, a

CC method of detecting a CA nucleic acid, a method of diagnosing cancer, a

CC method of treating cancer and a method of inhibiting expression of a CA

CC nucleic acid in a cell. The CA nucleic acids are useful for detecting CA

CC nucleic acids. The antibody is useful for detecting the presence or

CC absence of cancer cells in an individual which involves contacting cells

CC from the individual with the antibody and detecting a complex of a CA

CC protein from the cancer cells and the antibody, where the detection of

CC the complex correlates with the presence of cancer cells in the

CC individual. The composition is useful for inhibiting growth of cancer

CC cells in an individual or for delivering a therapeutic agent to cancer

CC cells in an individual. The invention is also useful for diagnosing

CC cancer, for treating cancer and for inhibiting expression of a CA gene in

CC a cell. This sequence represents murine cancer-associated genomic DNA of

XX the invention.

SQ Sequence 101636 BP; 28997 A; 21321 C; 20719 G; 29584 T; 0 U; 1015 Other;

Query Match 68.8%; Score 17.2; DB 14; Length 101636;

Best Local Similarity 86.4%; Pred. No. 1e+03; 3; Indels 0; Gaps 0;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCGTG 22

Db 25 CAAATTAAGAGACTATTCGTG 4

RESULT 58

AAV21209 10

Continuation (11 of 17) of AAV21209 from base 1000001 (Methanococcus jannaschii circular

WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209

WP Fragment Name Begin End

WP AAV21209\_00 1 110000

WP AAV21209\_01 100001 210000

WP AAV21209\_02 200001 310000

WP AAV21209\_03 300001 410000

WP AAV21209\_04 400001 510000

WP AAV21209\_05 500001 610000

WP AAV21209\_06 600001 710000

WP AAV21209\_07 700001 810000

WP AAV21209\_08 800001 910000

WP AAV21209\_09 900001 1010000

WP AAV21209\_10 1000001 1110000

WP AAV21209\_11 1100001 1210000

WP AAV21209\_12 1200001 1310000

WP AAV21209\_13 1300001 1410000

WP AAV21209\_14 1400001 1510000

WP AAV21209\_15 1500001 1610000

WP AAV21209\_16 1600001 1664976

Query Match 68.8%; Score 17.2; DB 2; Length 110000;

Best Local Similarity 86.4%; Pred. No. 1e+03;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 ATTAAGAGACTATTCGTGCAA 25

Db 49773 ATTAAGAGACTATTCGTGCAA 49794

RESULT 59

ACN43998\_5/c

Continuation (6 of 7) of ACN43998 from base 500001 (Human genomic sequence HCG1778483.)

WP Sequence split into 7 fragments LOCUS ACN43998 Accession Acn43998

WP Fragment Name Begin End

WP ACN43998\_0 1 110000

WP ACN43998\_1 100001 210000

WP ACN43998\_2 200001 310000

WP ACN43998\_3 300001 410000

WP ACN43998\_4 400001 510000

WP ACN43998\_5 500001 610000

WP ACN43998\_6 600001 653122

Query Match 68.8%; Score 17.2; DB 11; Length 110000;

Best Local Similarity 86.4%; Pred. No. 1e+03; Mismatches 3; Indels 0; Gaps 0;  
Matches 19; Conservative 0;

QY 1 CAAATTAAAGAGACTATTCGTG 22  
DB 107207 CAAATTAAAGAGAAATTTGTG 107186

RESULT 60

ACN45054  
ID ACN45054 standard; DNA; 133632 BP.

XX AC ACN45054;  
XX 18-NOV-2004 (first entry)  
XX Human genomic sequence hCG28560.  
XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.  
XX Homo sapiens.

XX WO2003073826-A2.  
XX 12-SEP-2003.  
XX 28-FEB-2003; 2003WO-US006235.  
XX 01-MAR-2002; 2002US-00087192.  
XX (SAGR-) SAGRES DISCOVERY.  
XX Morris DW;  
XX WPI; 2003-328604/31.  
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma  
XX comprises a nucleotide sequence.  
XX Claim 1; SEQ ID NO 1810; Opp; English.

XX The present invention relates to novel DNA and protein sequences which  
XX are associated with carcinomas. The sequences are useful for: (i) for  
XX screening drug candidates; (ii) for screening of bioactive agent capable  
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of  
XX a bioactive agent capable of modulating the activity of CAP; (iv) for  
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing  
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating  
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;  
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for  
XX determining Carcinoma Associated (CA) gene copy number. In addition, the  
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of  
XX carcinoma including lymphoma. The present sequence is one such CA coding  
XX sequence. Note: This patent is an equivalent to basic patent  
XX US2002182586A1, for which no sequence data was published

XX SQ Sequence 133632 BP; 39756 A; 25635 C; 26723 G; 41518 T; 0 U; 0 Other;

Query Match 68.8%; Score 17.2; DB 11; Length 133632;  
Best Local Similarity 86.4%; Pred. No. 1e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTCGTGCA 24  
DB 22221 AATAAGGAGACTATTCATGCA 22242

Search completed: January 27, 2006, 22:15:30  
Job time : 243.708 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 20:43:21 ; Search time 186.966 Seconds  
(without alignments)  
712.930 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaaatgcagaaacagg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

1: N\_Geneseq\_21.\*

2: Geneseqn1980s.\*

3: Geneseqn1990s.\*

4: Geneseqn2000s.\*

5: Geneseqn2001as.\*

6: Geneseqn2002as.\*

7: Geneseqn2002bs.\*

8: Geneseqn2003as.\*

9: Geneseqn2003bs.\*

10: Geneseqn2003cs.\*

11: Geneseqn2003ds.\*

12: Geneseqn2004as.\*

13: Geneseqn2004bs.\*

14: Geneseqn2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	14	ADZ75864
2	20	100.0	1731	6	ADZ75864 Group B s
3	20	100.0	1734	13	ADZ75864 Streptoc
4	20	100.0	29072	13	ADZ75864 Streptoc
5	20	100.0	29072	13	ADZ75864 Streptoc
6	20	100.0	110000	6	ADZ75864 Streptoc
7	20	100.0	110000	13	ADZ75864 Streptoc
8	18.4	92.0	9473	1	ADZ75864 Streptoc
9	17.4	87.0	272	4	ADZ75864 Streptoc
10	17.4	87.0	272	8	ADZ75864 Streptoc
11	17.4	87.0	732	8	ADZ75864 Streptoc
12	17.4	87.0	4911	4	ADZ75864 Streptoc
13	17.4	87.0	7886	4	ADZ75864 Streptoc
14	17.4	87.0	99090	12	ADZ75864 Streptoc
15	17.4	87.0	177587	11	ADZ75864 Streptoc
16	17.4	87.0	177930	14	ADZ75864 Streptoc
17	17	85.0	1081	2	ADZ75864 Streptoc
18	17	85.0	1081	6	ADZ75864 Streptoc
19	17	85.0	22467	4	ADZ75864 Streptoc

C	20	17	85.0	22467	4	AAS33423
C	21	16.8	84.0	310	2	AAQ77281
C	22	16.8	84.0	1119	3	AAC46080
C	23	16.8	84.0	2689	2	AAQ05230
C	24	16.8	84.0	5626	5	ABA19882
C	25	16.8	84.0	9360	2	AAQ02830
C	26	16.8	84.0	9630	3	AAZ89628
C	27	16.8	84.0	9633	1	AAZ89628
C	28	16.8	84.0	9633	1	AAZ89628
C	29	16.8	84.0	9633	1	AAZ89628
C	30	16.8	84.0	9670	12	ADL23697
C	31	16.8	84.0	9672	2	AAQ02616
C	32	16.8	84.0	9693	1	AAZ89628
C	33	16.8	84.0	9726	3	AAZ89628
C	34	16.8	84.0	9822	1	AAZ89628
C	35	16.8	84.0	14425	10	ADC87184
C	36	16.8	84.0	16270	10	ADC87444
C	37	16.8	84.0	41079	13	ABD33401
C	38	16.8	84.0	41104	14	ADZ13504
C	39	16.8	84.0	80268	13	ABD33401
C	40	16.4	82.0	453	2	AAT83759
C	41	16.4	82.0	469	13	ADQ55786
C	42	16.4	82.0	507	10	ABX08116
C	43	16.4	82.0	507	12	ADM91806
C	44	16.4	82.0	512	12	ACH75526
C	45	16.4	82.0	626	13	ACN52635
C	46	16.4	82.0	711	8	ACF73876
C	47	16.4	82.0	714	6	ABN93331
C	48	16.4	82.0	714	13	ADS04339
C	49	16.4	82.0	762	8	ACA24380
C	50	16.4	82.0	924	2	AAV75056
C	51	16.4	82.0	1035	2	AAV75056
C	52	16.4	82.0	1110	2	AAV75056
C	53	16.4	82.0	1196	3	AAA26350
C	54	16.4	82.0	1196	12	ADL71411
C	55	16.4	82.0	1638	2	AAV68066
C	56	16.4	82.0	1808	6	ABQ54797
C	57	16.4	82.0	2014	6	ABA90791
C	58	16.4	82.0	2019	6	ABA90848
C	59	16.4	82.0	3807	6	ABQ67188
C	60	16.4	82.0	3912	8	ACA22058

ALIGNMENTS

RESULT 1

ADZ75864

ID ADZ75864 standard; DNA; 20 BP.

XX AC ADZ75864;

XX DT 28-JUL-2005 (first entry)

XX DE Group B streptococcus phosphotransferase (pts) probe, SEQ ID NO:4.

XX DE Microorganism detection; fluorescence; diagnosis;

XX DE streptococcus infection; infection; gynecology and obstetrics;

XX DE phosphotransferase; probe; ss.

XX OS Streptococcus sp. 'group B'.

XX Key Location/Qualifiers

XX modified\_base 1 /\*tag= a

XX /mod\_base= OTHER

XX /note= "5" labeled with fluorescent acceptor LC-Red640"

XX US2005106578-A1.

XX 19-MAY-2005.

XX 18-NOV-2003; 2003US-00716005.

XX 18-NOV-2003; 2003US-00716005.  
PR (UHLJ/) UHL J R.  
XX (COCK/) COCKERILL F R.  
PA (AICH/) AICHINGER C.  
PA (REIS/) REISER A.  
XX  
PI Uhl JR, Cockerill FR, Aichinger C, Reiser A;  
XX WPI; 2005-371550/38.  
DR  
XX  
XX  
PT Detecting group B streptococcus, comprises amplifying a sample with pts  
PT primers, hybridizing the sample with fluorescently labeled pts probes,  
PT and detecting the presence of fluorescence resonance energy transfer.  
XX  
XX Claim 3; SEQ ID NO 4; 13pp; English.  
XX  
XX The invention relates to a real-time PCR-based method of detecting the  
CC presence or absence of group B streptococcus (GBS) bacterial pathogens in  
CC a biological sample from an individual. The method comprises  
CC amplification of a conserved region of the phosphotransferase (pts) gene  
CC of GBS (especially using PCR primers ADZ75861-ADZ75862), detection of the  
CC amplification product with a pair of fluorescently labeled  
CC phosphotransferase probes (especially ADZ75863-ADZ75864), and detection  
CC of the presence or absence of fluorescence resonance energy transfer  
CC (FRET), where presence of FRET is indicative of the presence of GBS in  
CC the sample. GBS infection is a leading cause of neonatal morbidity and  
CC mortality, with infection occurring during childbirth. Currently, it is  
CC recommended that women are screened for GBS during week 35-37 of  
CC gestation by a culture-based method which may take up to 72 hours for a  
CC result. However, many women first present at healthcare facilities at the  
CC time of labor, and in addition, GBS infection can be transient, so that a  
CC woman free of GBS at the time of screening may not be free of GBS when  
CC she is due to give birth. The method of the invention provides a real-  
CC time assay for the detection of group B streptococcus in a sample, is  
CC more sensitive and specific than prior art non-culture based methods, and  
CC can thus be implemented for routine diagnosis of the presence of group B  
CC streptococcus. Sequences ADZ75863-ADZ75864 represent fluorescently  
CC labeled GBS phosphotransferase probes specifically claimed for use in the  
CC method of the invention.  
XX  
XX Sequence 20 BP; 10 A; 3 C; 5 G; 2 T; 0 U; 0 Other;  
SQ  
Query Match 100.0%; Score 20; DB 14; Length 20;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAGG 20  
DB 1 CAAGTAAATGCAGAAACAGG 20  
RESULT 2  
ABN68276  
ID ABN68276 standard; DNA; 1731 BP.  
XX  
AC ABN68276;  
XX  
DT 01-JUL-2002 (first entry)  
XX  
DE Streptococcus polynucleotide SEQ ID NO 4465.  
XX  
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;  
KW antinflammatory; infection; vaccine; meningitis; gene therapy; ds.  
XX  
OS Streptococcus agalactiae.  
XX  
XX WO200234771-A2.  
XX  
PD 02-MAY-2002.  
XX

PF 29-OCT-2001; 2001WO-GB004789.  
XX  
PR 27-OCT-2000; 2000GB-00026333.  
PR 24-NOV-2000; 2000GB-00028727.  
PR 07-MAR-2001; 2001GB-00005640.  
XX  
PA (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;  
PI Tettelin H;  
XX  
XX WPI; 2002-352536/38.  
DR P-PSDB; ABP27645.  
DR  
XX  
PT New Streptococcus protein for the treatment or prevention of infection or  
PT disease caused by Streptococcus bacteria, such as meningitis, and for  
PT detecting a compound that binds to the protein.  
XX  
XX Claim 7; Page 3607; 4525pp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
SQ Sequence 1731 BP; 541 A; 323 C; 373 G; 494 T; 0 U; 0 Other;  
Query Match 100.0%; Score 20; DB 6; Length 1731;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAGG 20  
DB 292 CAAGTAAATGCAGAAACAGG 311  
RESULT 3  
ADV85417  
ID ADV85417 standard; DNA; 1734 BP.  
XX  
XX AC ADV85417;  
XX  
XX 24-FEB-2005 (first entry)  
XX  
XX Streptococcus agalactiae DNA sequence, SEQ ID 6558.  
DE  
XX Antibacterial; vaccine; bacterial infection; ds.  
XX  
OS Streptococcus agalactiae.  
XX  
XX WO200292818-A2.  
XX  
XX 21-NOV-2002.  
XX  
XX 26-APR-2002; 2002WO-IB003059.  
XX  
XX 26-APR-2001; 2001FR-00005642.  
XX  
XX (INSP ) INST PASTEUR.  
PA

PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.  
 XX  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 XX Claim 4; SEQ ID NO 6558; 439pp; French.  
 XX  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The  
 CC nucleotide sequences encode polypeptides of S. agalactiae involved in the  
 CC synthesis of amino acids, cell membranes, intermediate (central)  
 CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
 CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
 CC regulatory functions, replication, transcription, translation, protein  
 CC transport, adaptation to atypical conditions, sensitivity to medicines  
 CC and/or analogues, functions related to transposons, biosynthesis of  
 CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
 CC of nucleic acids. (I) are useful for the detection and/or amplification  
 CC useful for treatment of a bacterial S. agalactiae infection. The complete  
 CC genome of Streptococcus agalactiae is given in ADV81204. Note: The  
 CC present patent is an equivalent for the basic patent FR2824074A1, which  
 CC contains only 2344 sequences.  
 XX  
 XX Sequence 1734 BP; 540 A; 323 C; 376 G; 495 T; 0 U; 0 Other;  
 SQ  
 Query Match 100.0%; Score 20; DB 13; Length 1734;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACAGG 20  
 Db 292 CAAGTAAATGCAGAAACAGG 311  
 RESULT 4  
 ID ADV87723 standard; DNA; 23072 BP.  
 XX  
 XX ADV87723;  
 XX  
 XX 24-FEB-2005 (first entry)  
 XX  
 XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
 DE  
 XX Antibacterial; Vaccine; bacterial infection; ds.  
 XX  
 XX Streptococcus agalactiae.  
 XX  
 XX FR2824074-A1.  
 XX  
 XX 31-OCT-2002.  
 XX  
 XX 26-APR-2001; 2001FR-00005642.  
 XX  
 XX 26-APR-2001; 2001FR-00005642.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;  
 XX WPI; 2004-101891/11.  
 XX  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus

PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 XX Claim 1; SEQ ID NO 117; 2687pp; French.  
 XX  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
 CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
 CC agalactiae involved in the synthesis of amino acids, cell membranes,  
 CC intermediate (central) metabolism, energetic metabolism, fatty acid and  
 CC phospholipid metabolism, nucleotide metabolism including purines,  
 CC pyrimidines and/or nucleosides, regulatory functions, replication,  
 CC transcription, translation, protein transport, adaptation to atypical  
 CC conditions, sensitivity to medicines and/or analogues, functions related  
 CC to transposons, biosynthesis of cofactors, prosthetic groups and  
 CC transporters, cell membrane proteins and cellular machinery. (I) are  
 CC useful for the detection and/or amplification of nucleic acids.  
 CC Pharmaceutical composition comprising (I) or (II) are useful for  
 CC treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is  
 CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
 CC contains 6617 sequence whereas the present patent only contains 2344  
 CC sequences.  
 XX  
 XX Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;  
 SQ  
 Query Match 100.0%; Score 20; DB 13; Length 29072;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACAGG 20  
 Db 12187 CAAGTAAATGCAGAAACAGG 12206  
 RESULT 5  
 ADV78976  
 ID ADV78976 standard; DNA; 29072 BP.  
 XX  
 XX ADV78976;  
 XX  
 XX 24-FEB-2005 (first entry)  
 XX  
 XX Streptococcus agalactiae DNA sequence, SEQ ID 117.  
 DE  
 XX Antibacterial; vaccine; bacterial infection; ds.  
 XX  
 XX Streptococcus agalactiae.  
 XX  
 XX WO200292818-A2.  
 XX  
 XX 21-NOV-2002.  
 XX  
 XX 26-APR-2002; 2002WO-IB003059.  
 XX  
 XX 26-APR-2001; 2001FR-00005642.  
 XX  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
 PI Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
 XX WPI; 2004-101891/11.  
 XX  
 XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
 PT agalactiae for the development of vaccines, diagnostic tools, DNA chips  
 PT and identification of therapeutic targets.  
 XX  
 XX Claim 1; SEQ ID NO 117; 439pp; French.  
 XX  
 XX The present invention relates to novel Streptococcus agalactiae  
 CC nucleotide sequences (I; ADV78860-ADV78998 and ADV83341-ADV85476) and  
 CC novel polypeptides (II; ADV78999-ADV81203 and ADV81205-ADV83340). The

CC nucleotide sequences encode polypeptides of *S. agalactiae* involved in the  
CC synthesis of amino acids, cell membranes, intermediate (central)  
CC metabolism, energetic metabolism, fatty acid and phospholipid metabolism,  
CC nucleotide metabolism including purines, pyrimidines and/or nucleosides,  
CC regulatory functions, replication, transcription, translation, protein  
CC transport, adaptation to atypical conditions, sensitivity to medicines  
CC and/or analogues, functions related to transposons, biosynthesis of  
CC cofactors, prosthetic groups and transporters, cell membrane proteins and  
CC cellular machinery. (i) are useful for the detection and/or amplification  
CC of nucleic acids. Pharmaceutical composition comprising (i) or (ii) are  
CC useful for treatment of a bacterial *S. agalactiae* infection. The complete  
CC genome of *Streptococcus agalactiae* is given in ADV81204. Note: The  
CC present patent is an equivalent for the basic patent FR2824074A1, which  
CC contains only 2344 sequences.

XX  
SQ Sequence 29072 BP; 9616 A; 4551 C; 5572 G; 9333 T; 0 U; 0 Other;

Query Match 100.0%; Score 20; DB 13; Length 29072;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20

Db 12187 CAAGTAAATGCAGAAACAGG 12206

## RESULT 6

ABN71527\_07  
Continuation (8 of 22) of ABN71527 from base 700001 (Streptococcus polynucleotide SEQ ID  
WP Sequence split into 22 fragments LOCUS ABN71527 Accession Abn71527

Fragment Name	Begin	End
WP ABN71527_00	1	110000
WP ABN71527_01	100001	210000
WP ABN71527_02	200001	310000
WP ABN71527_03	300001	410000
WP ABN71527_04	400001	510000
WP ABN71527_05	500001	610000
WP ABN71527_06	600001	710000
WP ABN71527_07	700001	810000
WP ABN71527_08	800001	910000
WP ABN71527_09	900001	1010000
WP ABN71527_10	1000001	1110000
WP ABN71527_11	1100001	1210000
WP ABN71527_12	1200001	1310000
WP ABN71527_13	1300001	1410000
WP ABN71527_14	1400001	1510000
WP ABN71527_15	1500001	1610000
WP ABN71527_16	1600001	1710000
WP ABN71527_17	1700001	1810000
WP ABN71527_18	1800001	1910000
WP ABN71527_19	1900001	2010000
WP ABN71527_20	2000001	2110000
WP ABN71527_21	2100001	2155561

Query Match 100.0%; Score 20; DB 6; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20

Db 91057 CAAGTAAATGCAGAAACAGG 91076

## RESULT 7

ADV81204\_08  
Continuation (9 of 23) of ADV81204 from base 800001 (Streptococcus agalactiae complete  
WP Sequence split into 23 fragments LOCUS ADV81204 Accession Adv81204

Fragment Name	Begin	End
WP ADV81204_00	1	110000
WP ADV81204_01	100001	210000
WP ADV81204_02	200001	310000
WP ADV81204_03	300001	410000
WP ADV81204_04	400001	510000

WP	ADV81204_05	500001	610000
WP	ADV81204_06	600001	710000
WP	ADV81204_07	700001	810000
WP	ADV81204_08	800001	910000
WP	ADV81204_09	900001	1010000
WP	ADV81204_10	1000001	1110000
WP	ADV81204_11	1100001	1210000
WP	ADV81204_12	1200001	1310000
WP	ADV81204_13	1300001	1410000
WP	ADV81204_14	1400001	1510000
WP	ADV81204_15	1500001	1610000
WP	ADV81204_16	1600001	1710000
WP	ADV81204_17	1700001	1810000
WP	ADV81204_18	1800001	1910000
WP	ADV81204_19	1900001	2010000
WP	ADV81204_20	2000001	2110000
WP	ADV81204_21	2100001	2210000
WP	ADV81204_22	2200001	2217924

Query Match 100.0%; Score 20; DB 13; Length 110000;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAGG 20

Db 71392 CAAGTAAATGCAGAAACAGG 71411

## RESULT 8

AAAN92768

ID AAAN92768 standard; DNA; 9473 BP.

XX AC AAAN92768;

XX DT 24-OCT-2003 (revised)

XX DT 25-MAR-2003 (revised)

XX DT 18-MAY-1990 (first entry)

XX DE HIV-2 variant HIV-D194 clone.

XX KW HIV-2; AIDS; HIV-D194; ss.

XX OS Human immunodeficiency virus 2.

XX PN EP347365-A.

XX PD 20-DEC-1989.

XX PF 13-JUN-1989; 89EP-00710057.

XX PR 14-JUN-1988; 88DE-03820223.

XX PA (DIAG-) DIAGEN INST MOLEKUL.

XX PA (CHEM-) CHEMOTHERAPEUT GEORG-SPE.

XX PI Henco K, Vonbriesen H, Immelmann A, Kuhnel H, Dietrich U;

XX PI Rubsamenwa H, Adamski M;

XX DR WPI; 1989-372573/51.

XX PT New HIV-2 virus variant isolates - used to obtain prods. for identifying

XX PT infections by HIV-2 or for vaccines for protection against AIDS

XX PT infections.

XX PS Disclosure; Fig 4; 27pp; English.

XX CC The sequence characterises the clone HIV-D194, it starts with R/U5 region  
XX CC of the LTR and ends with the U5 region. The DNA can be used to raise Abs  
XX CC for serological detection, and polypeptides encoded by the sequence used  
XX CC in vaccination. (Updated on 25-MAR-2003 to correct PF field.) (Updated on  
XX CC 25-MAR-2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI  
XX CC field.) (Updated on 24-OCT-2003 to standardise OS field)



SQ Sequence 9473 BP; 3215 A; 1953 C; 2374 G; 1923 T; 0 U; 8 Other;  
Query Match 92.0%; Score 18.4; DB 1; Length 9473;  
Best Local Similarity 95.0%; Pred. No. 2.1e+02;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
DB 4213 CAAGTAAATGCAGAAATAGG 4232

RESULT 9  
AAS47567/c  
ID AAS47567 standard; DNA; 272 BP.  
XX  
AC AAS47567;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Enterococcus faecalis cellular proliferation inhibitory sequence #137.  
XX  
KW Antisense; ss; prokaryotic cellular proliferation; antibiotic;  
KW antibacterial; drug design.  
XX  
OS Enterococcus faecalis.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US009180.  
XX  
PR 21-MAR-2000; 2000US-0191078P.  
PR 23-MAY-2000; 2000US-0206848P.  
PR 26-MAY-2000; 2000US-0207727P.  
PR 23-OCT-2000; 2000US-0242578P.  
PR 27-NOV-2000; 2000US-0253625P.  
PR 22-DEC-2000; 2000US-0257931P.  
PR 16-FEB-2001; 2001US-0269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX  
XX WPI; 2001-611495/70.  
XX  
XX New polynucleotides for the identification and development of  
XX antibiotics, comprise sequences of antisense nucleic acids.  
XX  
PS Claim 1; SEQ ID NO 144; 51lpp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the genes,  
XX their use in the discovery of novel antibiotics, the essential genes  
XX themselves and the encoded proteins. The prokaryotes used are Escherichia  
XX coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
XX useful for the identification of potential new targets for antibiotic  
XX development. The antisense nucleic acids can also be used to identify  
XX proteins used in proliferation, to express these proteins, and to obtain  
XX antibodies capable of binding to the expressed proteins. The proteins can  
XX be used to screen compounds in rational drug discovery programmes. The  
XX antisense nucleic acid sequence is also useful to screen for homologous  
XX nucleic acids which are required for cell proliferation in a wide variety  
XX of organisms. The present sequence is an antisense oligonucleotide of the  
XX invention. Note: The sequence data for this patent did not form part of  
XX the printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 272 BP; 74 A; 60 C; 40 G; 98 T; 0 U; 0 Other;  
Query Match 87.0%; Score 17.4; DB 4; Length 272;

Best Local Similarity 94.7%; Pred. No. 4.3e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
DB 166 CAAGTAAATGAAGAAACAG 148

RESULT 10  
ACA12260/c  
ID ACA12260 standard; DNA; 272 BP.  
XX  
AC ACA12260;  
XX  
DT 27-OCT-2003 (revised)  
DT 19-JUN-2003 (first entry)  
XX  
DE Prokaryotic essential gene antisense oligonucleotide #130.  
XX  
KW Antisense; ss; prokaryotic essential gene; cell proliferation;  
KW drug design.  
XX  
OS Archaea.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 1; SEQ ID NO 130; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational



QY 2 AAGTAAATGCAGAAACAGG 20

XX Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss;  
KW  
XX  
OS Homo sapiens.  
XX  
PN WO2003073826-A2.

```
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
XX 01-MAR-2002; 2002US-00087192.
XX
XX (SAGR-) SAGRES DISCOVERY.
XX
XX Morris DW;
XX
XX WPI; 2003-328604/31.
XX
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
XX comprises a nucleotide sequence.
XX
XX Claim 1; SEQ ID NO 1438; Opp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
XX are associated with carcinomas. The sequences are useful for: (i) for
XX screening drug candidates; (ii) for screening of bioactive agent capable
XX of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
XX a bioactive agent capable of modulating the activity of CAP; (iv) for
XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
XX determining Carcinoma Associated (CA) gene copy number. In addition, the
XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
XX carcinoma including lymphoma. The present sequence is one such CA coding
XX sequence. Note: This patent is an equivalent to basic patent
XX US2002182586A1, for which no sequence data was published
XX
XX SQ Sequence 177587 BP; 49045 A; 38259 C; 39386 G; 50877 T; 0 U; 20 Other;
XX
XX Query Match 87.0%; Score 17.4; DB 11; Length 177587;
XX Best Local Similarity 94.7%; Pred. No. 7.5e+02;
XX Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 CAAGTAAATGCAGAAACAG 19
XX ||||| ||||| |||||
XX Db 102050 CAAGTAAATTCAGAAACAG 102068
XX
XX RESULT 16
XX ADZ13195
XX ID ADZ13195 standard; DNA; 177930 BP.
XX
XX AC ADZ13195;
XX
XX DT 16-JUN-2005 (first entry)
XX
XX DE Human cancer-associated genomic DNA #58.
XX
XX KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
XX cytostatic; gene; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO2005031001-A2.
XX
XX XX 07-APR-2005.
XX
XX PF 23-SEP-2004; 2004WO-US031617.
XX
XX XX 23-SEP-2003; 2003US-00669920.
XX
XX (CHIR ) CHIRON CORP.
XX
XX PI Morris DW, Malandro MS;
XX
XX WPI; 2005-273395/28.
XX
XX Nucleic acid array useful for detecting cancer associated nucleic acid,
XX
XX comprises two or more nucleic acid probes.
XX
XX Disclosure; SEQ ID NO 715; 198pp; English.
XX
XX The invention relates to a nucleic acid array for detecting a cancer
XX associated (CA) nucleic acid, comprising two or more nucleic acid probes.
XX The invention also relates to a peptide array comprising two or more
XX isolated polypeptides encoded by a CA nucleic acid sequence, a compound
XX that binds to a polypeptide, an isolated antibody or its fragment which
XX binds to a polypeptide, which is prepared by immunizing a host animal
XX with a composition comprising the polypeptide or its antigen binding
XX fragment and collecting cells from the host expressing antibodies against
XX the antigen or its antigen binding fragment, a composition comprising the
XX antibody and a carrier, a method of screening for anticancer activity, a
XX method of detecting a CA nucleic acid, a method of diagnosing cancer, a
XX nucleic acid in a cell. The CA nucleic acids are useful for detecting CA
XX absence of cancer cells in an individual which involves contacting cells
XX from the individual with the antibody and detecting a complex of a CA
XX protein from the cancer cells and the antibody, where the detection of
XX the complex correlates with the presence of cancer cells in the
XX individual. The composition is useful for inhibiting growth of cancer
XX cells in an individual or for delivering a therapeutic agent to cancer
XX cells in an individual. The invention is also useful for diagnosing
XX cancer, for treating cancer and for inhibiting expression of a CA gene in
XX a cell. This sequence represents human cancer-associated genomic DNA of
XX the invention.
XX
XX SQ Sequence 177930 BP; 49132 A; 38321 C; 39437 G; 51020 T; 0 U; 20 Other;
XX
XX Query Match 87.0%; Score 17.4; DB 14; Length 177930;
XX Best Local Similarity 94.7%; Pred. No. 7.5e+02;
XX Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX QY 1 CAAGTAAATGCAGAAACAG 19
XX ||||| ||||| |||||
XX Db 102050 CAAGTAAATTCAGAAACAG 102068
XX
XX RESULT 17
XX AAX13710/c
XX ID AAX13710 standard; DNA; 1081 BP.
XX
XX AC AAX13710;
XX
XX DT 19-MAR-1999 (first entry)
XX
XX DE Enterococcus faecalis genome contig SEQ ID NO:773.
XX
XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;
XX vaccine; attenuation; computer readable medium; ds.
XX
XX OS Enterococcus faecalis.
XX
XX PN WO9850555-A2.
XX
XX XX 12-NOV-1998.
XX
XX PF 04-MAY-1998; 98WO-US008985.
XX
XX XX 06-MAY-1997; 97US-0044031P.
XX
XX PR 16-MAY-1997; 97US-004655P.
XX
XX PR 14-NOV-1997; 97US-0066009P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Kunsch CA, Dillon PJ, Barash SC;
XX
XX WPI; 1999-045171/04.
XX
XX New isolated Enterococcus faecalis polynucleotides and polypeptides -
XX used to develop products for the detection of Enterococcus and for use in
```

PT vaccines for prevention or attenuation of Enterococcus infection.

XX Claim 1; Page 1979-1980; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC AAX12938 to AAX13919 represent these nucleotide sequences which are

CC primary nucleotide sequences, also known as contigs. The computer-based

CC system can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence of

CC Enterococcus faecalis in samples. They can also be used for diagnosing

CC Enterococcal infection in an animal and monitoring progression of

CC disease, and for identifying agents which can be used to modulate the

CC growth or pathogenicity of Enterococcus faecalis, or another related

CC organism, in vivo or in vitro. In particular the polypeptides encoded by

CC the Enterococcus faecalis nucleotide sequences can be used in vaccines to

XX prevent or attenuate an Enterococcal infection

SQ Sequence 1081 BP; 310 A; 237 C; 183 G; 334 T; 0 U; 17 Other;

Query Match 85.0%; Score 17; DB 2; Length 1081;

Best Local Similarity 89.5%; Pred. No. 7.3e+02;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CAAGTAAATGCAGAAACAG 19

Db 608 CAAGTAAATGCAGAAACAG 590

||||||| :|||||

RESULT 18

ABS99505/C ,

ID ABS99505 standard; DNA; 1081 BP.

XX ABS99505;

AC ABS99505;

XX 18-DEC-2002 (first entry)

DT Enterococcus faecalis contig sequence #773.

DE

XX Computer readable medium; Enterococcus faecalis; microbe; growth;

KW pathogenicity; vaccine; resistance; Enterococcal infection; commercial;

KW therapeutic; industrial; fermenting; sugar source; metabolite; vaccine;

KW biochip technology; antibacterial; modulator of nucleic acid expression;

KW contig; ds.

XX Enterococcus faecalis.

OS US2002120116-A1.

XX 29-AUG-2002.

PD 04-MAY-1998; 98US-00070927.

PF 04-MAY-1998; 98US-00070927.

XX (KUNS/) KUNSCH C A.

PA (DILL/) DILLON P J.

PA (BARA/) BARASH S.

XX Kunsch CA, Dillon PJ, Barash S;

PI WPI; 2002-750065/81.

XX Computer readable medium having recorded on it a Enterococcus faecalis

PT nucleotide sequence useful for detecting diseases related to Enterococcus

PT infections in animals.

XX Claim 1; Page; 119pp; English.

PS The present invention relates to a new computer readable medium with an

CC Enterococcus faecalis nucleotide sequence. The invention is useful to

CC diagnose the presence of E.faecalis in a sample or determining the

CC presence of a specific microbe in a sample. The invention is also useful

CC for modulating the growth or pathogenicity of E.faecalis, in a vaccine to

CC confer resistance to Enterococcal infection, for commercial, therapeutic

CC and industrial purposes, and for fermenting a particular sugar source or

CC to produce a particular metabolite. The invention is useful for detecting

CC diseases related to Enterococcus infections in animals, and for detecting

CC E.faecalis using biochip technology. The present nucleic acid sequence

CC represents an Enterococcus faecalis contig DNA sequence of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification but was obtained in electronic format directly from USPTO

CC at http:sequedata.uspto.gov

XX Sequence 1081 BP; 310 A; 237 C; 183 G; 334 T; 0 U; 17 Other;

SQ Query Match 85.0%; Score 17; DB 6; Length 1081;

Best Local Similarity 89.5%; Pred. No. 7.3e+02;

Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 CAAGTAAATGCAGAAACAG 19

Db 608 CAAGTAAATGCAGAAACAG 590

||||||| :|||||

RESULT 19

AAS33425

ID AAS33425 standard; DNA; 22467 BP.

XX AAS33425;

AC AAS33425;

XX 04-DEC-2001 (first entry)

DT DNA encoding human secreted protein, Seq ID No 708.

DE Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;

KW rheumatoid arthritis; antiarteriosclerotic; cardiant; vascular;

KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;

KW cytostatic; Alzheimer's disease; Parkinson's disease; human; cancer;

KW multiple sclerosis; cancer; hyperproliferative disorder; infection;

KW Gaucher's disease; neurological disease; cerebrovascular disorder;

KW thrombosis; wound healing; ds.

XX Homo sapiens.

OS WO200155326-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001347.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 05-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.

PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-451931/48.

New nucleic acids and polypeptides, useful for diagnosing, preventing or treating medical conditions.

Disclosure; SEQ ID NO 708; 753pp; English.

The invention relates to novel isolated nucleic acid molecules (I) encoding human secreted proteins (II). (I) and (II) are used to prevent, treat or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate expression of secreted proteins. (I) and complementary sequences may also be used as DNA probes in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate the presence of similar nucleic acid sequences in samples, and so which patients may be in need of restorative therapy. (II) may also be used as antigens in the production of antibodies and in assays to identify modulators (agonists and antagonists) of the expression and activity of the secreted proteins. The anti-(II) antibodies and antagonists may also be used to down regulate expression and activity of (II). The anti-(II) antibodies may also be used as diagnostic agents for detecting the presence of (II) in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). The disorders include for example: immune/autoimmune diseases (e.g. HIV (human immunodeficiency virus) infections, anaemia, rheumatoid arthritis and multiple sclerosis), cancers and hyperproliferative disorders (e.g. melanomas, neoplasms of the breast or liver, Sezary syndrome and Gaucher's disease), neurological diseases (e.g. Alzheimer's disease, Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and

CC thrombosis), infections caused by bacteria, viruses and fungi and ocular  
CC disorders (e.g. corneal infections). (i) and (ii), agonists, antagonists  
CC and antibodies can also be used to promote wound healing, maintain organs  
CC before transplantation, and support cell culture of primary tissues.  
CC AAS33043-AAS33486 represent human secreted protein coding sequences, PCR

Query Match 85.0%; Score 17; DB 4; Length 22467;  
Best Local Similarity 100.0%; Pred. No. 9.6e+02;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACA 18  
|||||  
Db 17162 AAGTAAATGCAGAAACA 17178

RESULT 20  
AAS33423/C  
ID AAS33423 standard; DNA; 22467 BP.

XX AAS33423;

XX 04-DEC-2001 (first entry)

DE DNA encoding human secreted protein, Seq ID No 706.

XX Immunomodulatory; human immunodeficiency virus; HIV; anaemia; angina;  
KW rheumatoid arthritis; antiarteriosclerotic; cardiac; vascular;  
KW cerebroprotective; thrombolytic; antimicrobial; ophthalmological;  
KW cytostatic; Alzheimer's disease; Parkinson's disease; cancer;  
KW multiple sclerosis; cancer; hyperproliferative disorder; infection;  
KW Gaucher's disease; neurological disease; cerebrovascular disorder;  
KW thrombosis; wound healing; ds.

XX Homo sapiens.

XX WO20015326-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001347.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.

PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-451931/48.	
XX		
PT	New nucleic acids and polypeptides, useful for diagnosing, preventing or	
PT	treating medical conditions.	
PS	Disclosure; SEQ ID NO 706; 753pp; English.	
XX		
CC	The invention relates to novel isolated nucleic acid molecules (I)	
CC	encoding human secreted proteins (II). (I) and (II) are used to prevent,	
CC	treat or ameliorate a medical condition in e.g. humans, mice, rabbits,	
CC	goats, horses, cats, dogs, chickens or sheep. (I) and (II) may be used in	
CC	the prevention, treatment and diagnosis of diseases associated with	
CC	inappropriate expression of secreted proteins. (I) and complementary	
CC	sequences may also be used as DNA probes in diagnostic assays (e.g.	
CC	polymerase chain reactions (PCR)) to detect and quantitate the presence	
CC	of similar nucleic acid sequences in samples, and so which patients may	
CC	be in need of restorative therapy. (II) may also be used as antigens in	
CC	the production of antibodies and in assays to identify modulators	
CC	(agonists and antagonists) of the expression and activity of the secreted	
CC	proteins. The anti-(II) antibodies and antagonists may also be used to	
CC	down regulate expression and activity of (II). The anti-(II) antibodies	
CC	may also be used as diagnostic agents for detecting the presence of (II)	
CC	in samples (e.g. by enzyme linked immunosorbent assay (ELISA)). The	
CC	disorders include for example: immune/autoimmune diseases (e.g. HIV	
CC	(human immunodeficiency virus) infections, anaemia, rheumatoid arthritis	
CC	and multiple sclerosis), cancers and hyperproliferative disorders (e.g.	
CC	melanomas, neoplasms of the breast or liver, Sezary syndrome and	
CC	Gaucher's disease), neurological diseases (e.g. Alzheimer's disease,	
CC	Parkinson's disease and Charcot-Marie-Tooth disease), cardio-/	
CC	cerebrovascular disorders (e.g. cardiac arrest, tachycardia, angina and	
CC	thrombosis), infections caused by bacteria, viruses and fungi and ocular	
CC	disorders (e.g. corneal infections). (I) and (II), agonists, antagonists	
CC	and antibodies can also be used to promote wound healing, maintain organs	
CC	before transplantation, and support cell culture of primary tissues. PCR	
CC	AA533043-AA533486 represent human secreted protein coding sequences. PCR	
Query Match	85.0%;	Score 17; DB 4; Length 22467;



protein identification; signal transduction pathway; metabolic pathway;  
promoter; termination sequence; ss.

KW	Arabidopsis thaliana.	PR	24-JUN-1999;	99US-0140695P.
XX		PR	28-JUN-1999;	99US-0140823P.
XX		PR	29-JUN-1999;	99US-0140991P.
OS		PR	30-JUN-1999;	99US-0141287P.
XX		PR	01-JUL-1999;	99US-0141842P.
PN		PR	01-JUL-1999;	99US-0142154P.
PN		PR	02-JUL-1999;	99US-0142055P.
PD		PR	06-JUL-1999;	99US-0142390P.
XX		PR	08-JUL-1999;	99US-0142803P.
XX		PR	09-JUL-1999;	99US-0142920P.
PF		PR	12-JUL-1999;	99US-0142977P.
XX		PR	13-JUL-1999;	99US-0143542P.
XX		PR	14-JUL-1999;	99US-0143624P.
PR		PR	15-JUL-1999;	99US-0144005P.
PR		PR	16-JUL-1999;	99US-0144085P.
PR		PR	16-JUL-1999;	99US-0144086P.
PR		PR	19-JUL-1999;	99US-0144325P.
PR		PR	19-JUL-1999;	99US-0144331P.
PR		PR	19-JUL-1999;	99US-0144332P.
PR		PR	19-JUL-1999;	99US-0144333P.
PR		PR	19-JUL-1999;	99US-0144334P.
PR		PR	19-JUL-1999;	99US-0144335P.
PR		PR	20-JUL-1999;	99US-0144352P.
PR		PR	20-JUL-1999;	99US-0144632P.
PR		PR	20-JUL-1999;	99US-0144884P.
PR		PR	21-JUL-1999;	99US-0144814P.
PR		PR	21-JUL-1999;	99US-0145086P.
PR		PR	21-JUL-1999;	99US-0145088P.
PR		PR	22-JUL-1999;	99US-0145085P.
PR		PR	22-JUL-1999;	99US-0145087P.
PR		PR	22-JUL-1999;	99US-0145089P.
PR		PR	22-JUL-1999;	99US-0145192P.
PR		PR	23-JUL-1999;	99US-0145145P.
PR		PR	23-JUL-1999;	99US-0145218P.
PR		PR	23-JUL-1999;	99US-0145224P.
PR		PR	26-JUL-1999;	99US-0145276P.
PR		PR	27-JUL-1999;	99US-0145913P.
PR		PR	27-JUL-1999;	99US-0145918P.
PR		PR	27-JUL-1999;	99US-0145919P.
PR		PR	28-JUL-1999;	99US-0145951P.
PR		PR	02-AUG-1999;	99US-0146386P.
PR		PR	02-AUG-1999;	99US-0146388P.
PR		PR	03-AUG-1999;	99US-0146389P.
PR		PR	03-AUG-1999;	99US-0147038P.
PR		PR	04-AUG-1999;	99US-0147204P.
PR		PR	04-AUG-1999;	99US-0147302P.
PR		PR	05-AUG-1999;	99US-0147192P.
PR		PR	05-AUG-1999;	99US-0147260P.
PR		PR	06-AUG-1999;	99US-0147303P.
PR		PR	06-AUG-1999;	99US-0147416P.
PR		PR	09-AUG-1999;	99US-0147493P.
PR		PR	09-AUG-1999;	99US-0147935P.
PR		PR	10-AUG-1999;	99US-0148171P.
PR		PR	11-AUG-1999;	99US-0148319P.
PR		PR	12-AUG-1999;	99US-0148341P.
PR		PR	13-AUG-1999;	99US-0148565P.
PR		PR	13-AUG-1999;	99US-0148684P.
PR		PR	16-AUG-1999;	99US-0149368P.
PR		PR	17-AUG-1999;	99US-0149175P.
PR		PR	18-AUG-1999;	99US-0149426P.
PR		PR	20-AUG-1999;	99US-0149722P.
PR		PR	20-AUG-1999;	99US-0149723P.
PR		PR	20-AUG-1999;	99US-0149929P.
PR		PR	23-AUG-1999;	99US-0149902P.
PR		PR	23-AUG-1999;	99US-0149930P.
PR		PR	25-AUG-1999;	99US-0150566P.
PR		PR	26-AUG-1999;	99US-0150884P.
PR		PR	27-AUG-1999;	99US-0151065P.
PR		PR	27-AUG-1999;	99US-0151066P.
PR		PR	30-AUG-1999;	99US-0151080P.
PR		PR	31-AUG-1999;	99US-0151303P.
PR		PR	01-SEP-1999;	99US-0151438P.
PR		PR		99US-0151930P.

```
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      84.0%; Score 16.8; DB 3; Length 1119;
Best Local Similarity 90.0%; Pred. No. 9e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
DB 244 CAAGTAAATGCAGAAATCATG 225

RESULT 23
AAX05230
ID AAX05230 standard; DNA; 2689 BP.
XX
XX AAX05230;
XX
XX 17-OCT-2003 (revised)
XX 22-APR-1999 (first entry)
XX
XX HIV-2 genomic DNA sequence.
XX
XX HIV-1; HIV-2; Human immunodeficiency virus; primer/probe set; HIV;
XX detection; nucleic acid amplification; PCR; RT PCR; OH PCR; virus;
XX oligonucleotide hybridisation PCR; ss.
XX
XX Human immunodeficiency virus 2.
XX
XX WO9858086-A2.
```

---

```
XX 23-DEC-1998.
PD
XX
XX 04-JUN-1998; 98WO-US011652.
XX
XX 16-JUN-1997; 97US-00876546.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Kroeger PE, Abravaya K, Esping CAC, Gorzowski JJ, Hoenle RJ;
XX Moore JJ;
XX
XX WPI; 1999-095352/08.
XX
XX Nucleic acid primers and probes for detecting HIV 1 and HIV 2 - using
XX known nucleic acid amplification procedures especially oligonucleotide
XX hybridisation PCR.
XX
XX Disclosure; Page 27-28; 35pp; English.
XX
XX The invention provides primer/probe sets for detecting HIV-1 or HIV-2
XX that comprise two primers and at least one probe. The primer/probe sets
XX are useful to detect HIV (i.e. HIV-1 and HIV-2, either separately or
XX simultaneously) in biological samples using known nucleic acid
XX amplification procedures e.g. PCR, reverse transcriptase (RT) PCR and
XX especially oligonucleotide hybridisation PCR (OH PCR); HIV-1 and HIV-2
XX are different viruses each with several subtypes due to the highly
XX mutable nature of the virus (attributed to the inefficiency with which it
XX converts its genetic material (RNA) into DNA to allow it to insert its
XX genetic information into the host and/or recombination of viral genomes
XX from different HIV populations). The sequences are designed to detect all
XX known HIV-1 or HIV-2 subtypes by OH PCR and allow detection of target
XX sequences which are DNA or, sequences which are embedded within the HIV
XX genome and are therefore RNA by RT-PCR. Sequences AAX05232 to AAX05253
XX represent specifically claimed primer and probe sequences present in
XX eight different primer/probe sets that are used for detecting HIV-1 and
XX HIV-2. The present sequence represents HIV-1 genomic DNA sequence that
XX can be used as a target sequence in the method of the invention. (Updated
XX on 17-OCT-2003 to standardise OS field)
XX
XX Sequence 2689 BP; 1066 A; 487 C; 597 G; 539 T; 0 U; 0 Other;
Query Match      84.0%; Score 16.8; DB 2; Length 2689;
Best Local Similarity 90.0%; Pred. No. 9.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
DB 2139 CAAGTAAATGCAGAACTAGG 2158

RESULT 24
ABA19882
ID ABA19882 standard; DNA; 5626 BP.
XX
XX ABA19882;
XX
XX 23-JAN-2002 (first entry)
XX
XX Human nervous system related polynucleotide SEQ ID NO 12213.
XX
XX Human; neutropic; neuroprotective; cytostatic; dermatological; virucide;
XX immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnary;
XX antiparkinsonian; antiskilling; antianaemic; antirheumatic; cancer;
XX antirheumatic; hepatotropic; cerebrotropic; antineoplastic;
XX antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX antiparasitic; cardiant; immune disorder; cardiovascular disorder;
XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX
XX WO200159063-A2.
XX
```

PD	16-AUG-2001.	PR	29-SEP-2000; 2000US-0236368P.
XX		PR	29-SEP-2000; 2000US-0236369P.
PF	17-JAN-2001; 2001WO-US001334.	PR	29-SEP-2000; 2000US-0236370P.
XX		PR	02-OCT-2000; 2000US-0236802P.
PR	31-JAN-2000; 2000US-0179065P.	PR	02-OCT-2000; 2000US-0237038P.
PR	04-FEB-2000; 2000US-0180628P.	PR	02-OCT-2000; 2000US-0237039P.
PR	24-FEB-2000; 2000US-0184664P.	PR	02-OCT-2000; 2000US-0237040P.
PR	02-MAR-2000; 2000US-0186350P.	PR	13-OCT-2000; 2000US-0239935P.
PR	16-MAR-2000; 2000US-0189974P.	PR	13-OCT-2000; 2000US-0239937P.
PR	17-MAR-2000; 2000US-0190076P.	PR	20-OCT-2000; 2000US-0240960P.
PR	18-APR-2000; 2000US-0198123P.	PR	20-OCT-2000; 2000US-0241785P.
PR	19-MAY-2000; 2000US-0205515P.	PR	20-OCT-2000; 2000US-0241786P.
PR	07-JUN-2000; 2000US-0209467P.	PR	20-OCT-2000; 2000US-0241808P.
PR	28-JUN-2000; 2000US-0214886P.	PR	20-OCT-2000; 2000US-0241809P.
PR	30-JUN-2000; 2000US-0215135P.	PR	20-OCT-2000; 2000US-0241826P.
PR	07-JUL-2000; 2000US-0216647P.	PR	20-OCT-2000; 2000US-0242221P.
PR	07-JUL-2000; 2000US-0216880P.	PR	01-NOV-2000; 2000US-0244617P.
PR	11-JUL-2000; 2000US-0217487P.	PR	08-NOV-2000; 2000US-0246474P.
PR	14-JUL-2000; 2000US-0218290P.	PR	08-NOV-2000; 2000US-0246475P.
PR	26-JUL-2000; 2000US-0220963P.	PR	08-NOV-2000; 2000US-0246477P.
PR	26-JUL-2000; 2000US-0220964P.	PR	08-NOV-2000; 2000US-0246478P.
PR	14-AUG-2000; 2000US-0224518P.	PR	08-NOV-2000; 2000US-0246523P.
PR	14-AUG-2000; 2000US-0224519P.	PR	08-NOV-2000; 2000US-0246524P.
PR	14-AUG-2000; 2000US-022513P.	PR	08-NOV-2000; 2000US-0246525P.
PR	14-AUG-2000; 2000US-0225214P.	PR	08-NOV-2000; 2000US-0246526P.
PR	14-AUG-2000; 2000US-0225266P.	PR	08-NOV-2000; 2000US-0246527P.
PR	14-AUG-2000; 2000US-0225267P.	PR	08-NOV-2000; 2000US-0246528P.
PR	14-AUG-2000; 2000US-0225447P.	PR	08-NOV-2000; 2000US-0246532P.
PR	14-AUG-2000; 2000US-0225757P.	PR	08-NOV-2000; 2000US-0246609P.
PR	14-AUG-2000; 2000US-0225758P.	PR	08-NOV-2000; 2000US-0246610P.
PR	14-AUG-2000; 2000US-0225759P.	PR	08-NOV-2000; 2000US-0246611P.
PR	18-AUG-2000; 2000US-0226681P.	PR	08-NOV-2000; 2000US-0246613P.
PR	22-AUG-2000; 2000US-0226868P.	PR	17-NOV-2000; 2000US-0249207P.
PR	22-AUG-2000; 2000US-0227182P.	PR	17-NOV-2000; 2000US-0249208P.
PR	23-AUG-2000; 2000US-0227009P.	PR	17-NOV-2000; 2000US-0249209P.
PR	30-AUG-2000; 2000US-0228924P.	PR	17-NOV-2000; 2000US-0249210P.
PR	01-SEP-2000; 2000US-0229343P.	PR	17-NOV-2000; 2000US-0249211P.
PR	01-SEP-2000; 2000US-0229347P.	PR	17-NOV-2000; 2000US-0249212P.
PR	01-SEP-2000; 2000US-0229344P.	PR	17-NOV-2000; 2000US-0249213P.
PR	01-SEP-2000; 2000US-0229345P.	PR	17-NOV-2000; 2000US-0249214P.
PR	05-SEP-2000; 2000US-0229509P.	PR	17-NOV-2000; 2000US-0249215P.
PR	06-SEP-2000; 2000US-0229513P.	PR	17-NOV-2000; 2000US-0249216P.
PR	06-SEP-2000; 2000US-0230437P.	PR	17-NOV-2000; 2000US-0249217P.
PR	08-SEP-2000; 2000US-0230438P.	PR	17-NOV-2000; 2000US-0249218P.
PR	08-SEP-2000; 2000US-0231242P.	PR	17-NOV-2000; 2000US-0249219P.
PR	08-SEP-2000; 2000US-0231243P.	PR	17-NOV-2000; 2000US-0249244P.
PR	08-SEP-2000; 2000US-0231244P.	PR	17-NOV-2000; 2000US-0249245P.
PR	08-SEP-2000; 2000US-0231413P.	PR	17-NOV-2000; 2000US-0249264P.
PR	08-SEP-2000; 2000US-0231414P.	PR	17-NOV-2000; 2000US-0249265P.
PR	08-SEP-2000; 2000US-0232080P.	PR	17-NOV-2000; 2000US-0249297P.
PR	08-SEP-2000; 2000US-0232081P.	PR	17-NOV-2000; 2000US-0249299P.
PR	12-SEP-2000; 2000US-0231968P.	PR	17-NOV-2000; 2000US-0249300P.
PR	14-SEP-2000; 2000US-0232397P.	PR	01-DEC-2000; 2000US-0250391P.
PR	14-SEP-2000; 2000US-0232398P.	PR	01-DEC-2000; 2000US-0251160P.
PR	14-SEP-2000; 2000US-0232399P.	PR	05-DEC-2000; 2000US-0251030P.
PR	14-SEP-2000; 2000US-0232400P.	PR	05-DEC-2000; 2000US-0251988P.
PR	14-SEP-2000; 2000US-0232401P.	PR	05-DEC-2000; 2000US-0256719P.
PR	14-SEP-2000; 2000US-0233063P.	PR	06-DEC-2000; 2000US-0251479P.
PR	14-SEP-2000; 2000US-0233064P.	PR	08-DEC-2000; 2000US-0251856P.
PR	14-SEP-2000; 2000US-0233065P.	PR	08-DEC-2000; 2000US-0251868P.
PR	21-SEP-2000; 2000US-0234223P.	PR	08-DEC-2000; 2000US-0251869P.
PR	21-SEP-2000; 2000US-0234274P.	PR	08-DEC-2000; 2000US-0251989P.
PR	25-SEP-2000; 2000US-0234997P.	PR	08-DEC-2000; 2000US-0251990P.
PR	25-SEP-2000; 2000US-0234998P.	PR	11-DEC-2000; 2000US-0254097P.
PR	26-SEP-2000; 2000US-0235484P.	PR	05-JAN-2001; 2001US-0259678P.
PR	27-SEP-2000; 2000US-0235834P.	XX	(HUMA-) HUMAN GENOME SCI INC.
PR	27-SEP-2000; 2000US-0235836P.	XX	Rosen CA, Barash SC, Ruben SM;
PR	29-SEP-2000; 2000US-0236327P.	XX	WPI; 2001-541565/60.
PR	29-SEP-2000; 2000US-0236367P.	XX	
		DR	

XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system cancers  
PT and metastases.  
XX  
PS Disclosure; SEQ ID NO 12213; 1701pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (ABP1478-ABP18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 5626 BP; 1311 A; 1482 C; 1531 G; 1302 T; 0 U; 0 Other;  
Query Match 84.0%; Score 16.8; DB 5; Length 5626;  
Best Local Similarity 90.0%; Pred. No. 1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAGG 20  
DB 5588 CATGTAATGCAGAAACATG 5607  
RESULT 25  
AAQ02830  
ID AAQ02830 standard; cDNA; 9360 BP.  
XX  
AC AAQ02830;  
XX  
DT 24-OCT-2003 (revised)  
DT 25-MAR-2003 (revised)  
DT 29-MAY-1989 (first entry)  
XX  
DE cDNA to HIV-2 RNA.  
XX  
KW HIV; AIDS; Vaccine; pUC-HIV-2 (GH-1).  
XX  
OS Human immunodeficiency virus 2.  
XX  
Key Location/Qualifiers  
FH 544..2109  
CDS /tag= a  
FT /label= Gag reading frame  
FT 1830..4934  
FT /tag= b  
FT /label= Pol gene reading frame.  
FT 4867..5511  
FT /tag= c  
FT /label= Q gene reading frame  
FT 5342..5677  
FT /tag= d  
FT /label= X gene reading frame  
FT 5680..5994  
FT /tag= e  
FT /label= R gene reading frame  
FT 6148..8703  
FT /tag= f  
FT /label= Env gene reading frame.  
FT 8540..9304  
FT /tag= g  
FT

FT /label= F gene reading frame.  
XX JP01289486-A.  
PN 21-NOV-1989.  
XX  
PF 16-MAY-1988; 88JP-00119024.  
XX  
PR 16-MAY-1988; 88JP-00119024.  
XX  
PA (TOFU ) TOA NENRYO KOGYO KK.  
PA (FJRE ) FUJI REBIO KK.  
XX  
DR WPI: 1990-005177/01.  
DR P-FSDB; AAR04024, AAR04025, AAR04026, AAR04027, AAR04028, AAR04029,  
DR AAR04030.  
XX  
PT DNA indicating complement to RNA gene - of human immuno: deficiency virus  
PT type 2. used for new vaccine or diagnostic for aids virus.  
XX  
PS Claim 2; Fig 4; 12pp; Japanese.  
XX  
CC cDNA to novel HIV-2 (GH-1) has been integrated into plasmid pUC HIV-2 (GH-  
CC 1). Useful for diagnosis and vaccination against the virus. Carries 7  
CC overlapping genes in various reading frames, including Gag, Pol and  
CC Env. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-  
CC 2003 to standardise OS field)  
XX  
SQ Sequence 9360 BP; 3211 A; 1944 C; 2317 G; 1888 T; 0 U; 0 Other;  
Query Match 84.0%; Score 16.8; DB 2; Length 9360;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAGG 20  
DB 4212 CAGGTAATGCAGAAATAGG 4231  
RESULT 26  
AAZ89628  
ID AAZ89628 standard; DNA; 9630 BP.  
XX  
AC AAZ89628;  
XX  
DT 15-SEP-2003 (revised)  
DT 11-JUL-2000 (first entry)  
XX  
DE HIV-2 ROD DNA.  
XX  
KW Diagnosis; probe; lymphocyte; virus; immunodiagnostic; infection;  
KW antiviral; ss.  
XX  
OS Human immunodeficiency virus 2.  
XX  
PN US6054565-A.  
XX  
PD 25-APR-2000.  
XX  
PF 28-APR-1994; 94US-00234875.  
XX  
PR 03-MAR-1986; 86US-00835228.  
PR 06-OCT-1986; 86US-00916080.  
PR 21-NOV-1986; 86US-00933184.  
PR 16-JAN-1987; 87US-00003764.  
PR 11-FEB-1987; 87US-00013477.  
PR 03-SEP-1991; 91US-00752368.  
PR 20-DEC-1991; 91US-00810908.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PI Montagnier L, Clavel F, Guyader M, Geutard D, Sonigo P, Alizon M;  
XX

DR WPI; 2000-328365/28.  
 XX Novel cloned nucleotide sequences homologous or identical to the portion  
 PT of genomic RNA of HIV-2 viruses useful as probes and in diagnostic tests  
 PT to diagnose HIV-2 infection.  
 XX  
 XX Claim 1; Col 10-23; 33pp; English.  
 XX  
 XX This invention describes a novel cloned nucleic acid (I) of a human  
 CC immunodeficiency virus type 2 (HIV-2). (I) is capable of being used as  
 CC probes in diagnostic method to obtain the immunological reagents  
 CC necessary to diagnose an HIV-2 infection. These sequences may be used as  
 CC probes in hybridization reactions with the genetic material of infected  
 CC patients to indicate whether the RNA of the HIV-2 virus is present in  
 CC these patient's lymphocytes or whether an analogous DNA is present. The  
 CC genetic sequence of the HIV-2 virus may be used to create the  
 CC polypeptides encoded by these sequences. Specifically, these polypeptides  
 CC may be created by expression of the cDNA obtained from bacterial, yeast  
 CC or animal cells. These polypeptides may be used in diagnostic tests such  
 CC as immunofluorescence assays, radioimmunoassays (RIA) and Western Blot  
 CC tests. Monoclonal antibodies to these polypeptides of fragments may be  
 CC created and used in immunodiagnostic tests. The polypeptides of the  
 CC present invention may also be used as immunogenic reagents to induce  
 CC protection against infection by HIV-2 viruses. The polypeptides produced  
 CC by recombinant-DNA techniques would function as vaccine agents. The  
 CC polypeptides may be used on competitive assays to test the ability of  
 CC various antiviral agents to determined their ability to prevent the virus  
 CC from fixing on its target. This sequence encodes the HIV-2 ROD genome  
 CC which is described in the method of the invention. (Updated on 15-SEP-  
 CC 2003 to standardise OS field)  
 XX  
 SQ Sequence 9630 BP; 3310 A; 1959 C; 2373 G; 1976 T; 0 U; 12 Other;  
 Query Match 84.0%; Score 16.8; DB 3; Length 9630;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACAGG 20  
 |||||  
 DB 4213 CAAGTAAATGCAGAACTAGG 4232  
 |||||  
 RESULT 27  
 AAN80890  
 ID AAN80890 standard; cDNA; 9633 BP.  
 XX  
 AC AAN80890;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 07-NOV-1990 (first entry)  
 XX  
 DE Sequence of cDNA clone HIV-2 SBL/ISY of HIV related retrovirus strain.  
 XX  
 KW HIV vaccine; HIV strain SBL-6669-85; ss.  
 XX  
 OS Human immunodeficiency virus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..9633  
 FT /\*tag= a  
 FT /note= "SQ in AAP81769"  
 FT CDS 2..9631  
 FT /\*tag= b  
 FT /note= "SQ in AAP81770"  
 FT CDS 3..9632  
 FT /\*tag= c  
 FT /note= "SQ in AAP81771"  
 XX  
 XX WO8808449-A.  
 XX  
 XX 03-NOV-1988.  
 XX  
 XX 28-APR-1988; 88WO-SE000218.  
 PF

XX 28-APR-1987; 87SE-00001765.  
 XX  
 XX (SBLs-) SBL STATENS BAKTERI.  
 PA (STAT-) STATENS BAKTERIOLOGISKA LAB.  
 PA (STAT-) STATENS BAKTERIOLOGISKA LAB.  
 XX  
 XX Albert J, Biberfeld G, Fenyo EM, Norrby E;  
 XX  
 XX WPI; 1988-322769/45.  
 DR P-PSDB; AAP81769, AAP81770, AAP81771.  
 DR  
 XX HIV related human retro-virus strain - used for obtaining antigens for  
 PT assays and vaccines and for prodn. of antibodies for assays.  
 PT  
 XX Claim 7; Fig 4; 28pp; English.  
 XX  
 XX cDNA clone (HIV-2 SBL/ISY) characterised by having the primary nucleotide  
 CC sequence in n80890 or a degenerate or a part thereof, esp. a clone having  
 CC capacity to give prodn. of infectious virus particles is claimed. HIV-2  
 CC SBL/ISY represents the complete genome of the virus SBL-6699 (=SBL-6689-  
 CC 85). The proviral DNA was obtd. from a genomic library constructed from  
 CC HUT-78 cells infected with SBL-6669-85 virus using the lambda-phage  
 CC vector EMBL-3. SBL-6669-85 was isolated from lymphocytes of a West  
 CC African woman. Protection is requested for the entire genome disclosed in  
 CC n80890 and for parts thereof, and corresp. to various genes such as the  
 CC gag gene (corresp. to nucleotides 547 to 2106), the pol gene (nucleotides  
 CC 1827-4931) and the env gene (nucleotides 6144 to 8682), the corresp. AA  
 CC sequences and parts thereof and various products derived therefrom, or  
 CC use thereof, such as clones prepd. by recombinant vector method, HIV test  
 CC devices and methods. (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 9633 BP; 3288 A; 1943 C; 2394 G; 2008 T; 0 U; 0 Other;  
 Query Match 84.0%; Score 16.8; DB 1; Length 9633;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACAGG 20  
 |||||  
 DB 4209 CAAGTAAATGCAGAACTAGG 4228  
 |||||  
 RESULT 28  
 AAN80859  
 ID AAN80859 standard; cDNA; 9643 BP.  
 XX  
 AC AAN80859;  
 XX  
 DT 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 15-NOV-1990 (first entry)  
 XX  
 DE Sequence of entire HIV-2 ROD genome.  
 XX  
 KW LAV-II ROD; ss.  
 XX  
 OS Human immunodeficiency virus; ROD.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 545..2110  
 FT /\*tag= a  
 FT /note= "AAP80801"  
 FT CDS 1828..4908  
 FT /\*tag= b  
 FT /note= "AAP81773"  
 FT CDS 4841..5485  
 FT /\*tag= c  
 FT /note= "AAP81774"  
 FT CDS 5316..5654  
 FT /\*tag= d  
 FT /note= "AAP81775"  
 FT CDS 5654..5968

```

FT      /*tag= e
FT      /note= "AAP81776"
FT      CDS      5817..6113
FT      /*tag= f
FT      /note= "AAP81777"
FT      CDS      6043..6153
FT      /*tag= g
FT      /note= "AAP81778"
FT      CDS      6119..8692
FT      /*tag= h
FT      /note= "AAP81779"
FT      CDS      8220..8372
FT      /*tag= i
FT      /note= "AAP81780"
FT      CDS      8242..8508
FT      /*tag= j
FT      /note= "AAP81781"
FT      CDS      8529..9299
FT      /*tag= k
FT      /note= "AAP81782"
FT      CDS
FT      WO8805440-A.
FT      PN
XX      XX
XX      28-JUL-1988.
XX      PD
XX      15-JAN-1988; 88WO-BF000025.
XX      PR      16-JAN-1987; 87US-00003764.
XX      PR      11-FEB-1987; 87FR-00001739.
XX      PR      15-APR-1987; 87FR-00005398.
XX      XX
XX      (INSP ) INST PASTEUR.
XX      PA      (ALIZ/) ALIZON M.
XX      PA      (CNRS ) CENT NAT RECH SCI.
XX      XX
XX      Montagnier L, Guetard D, Clavel F, Sonigo P, Guyader M;
XX      PI      Tiollais P, Chakrabar L;
XX      PI
XX      WPI; 1988-220290/31.
XX      DR      P-PSDB; AAP80801, AAP81773, AAP81774, AAP81775, AAP81776, AAP81777,
XX      DR      AAP81778, AAP81779, AAP81780, AAP81781, AAP81782.
XX      XX
XX      NEW peptide(s) with immunological properties of HIV-2 envelope protein -
XX      PT      have the structure of simian immune deficiency virus proteins, useful in
XX      PT      diagnosis and of vaccine components.
XX      XX
XX      Disclosure; Fig 1A; 86pp; French.
XX      PS
XX      The SQ in AAN80859 was deposited on 21/2/86 at the CNCM under number I-
XX      CC      522, reference name LAV-11 ROD. It is the cDNA to HIV-2 ROD genomic RNA.
XX      CC      SQ was compared with the SQ of the genome of SIV (Mac) (AAN80860) to
XX      CC      identify common regions. (Updated on 25-MAR-2003 to correct PF field.)
XX      CC      (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
XX      CC      correct PA field.) (Updated on 24-OCT-2003 to standardise OS field)
XX      XX
XX      Sequence 9643 BP; 3296 A; 1969 C; 2399 G; 1979 T; 0 U; 0 Other;
XX      SQ      Query Match      84.0%; Score 16.8; DB 1; Length 9643;
XX      Best Local Similarity 90.0%; Pred. No. 1.1e+03;
XX      Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX      QY      1 CAAGTAAATGCAGAACAGG 20
XX      Db      4213 CAAGTAAATGCAGAACACTAGG 4232
XX      RESULT 29
XX      AAC81938
XX      ID      AAC81938 standard; DNA; 9663 BP.
XX      AC
XX      AAC81938;
XX      XX
XX      28-FEB-2001 (first entry)
XX      DT

XX      Packaging vector PCM-ROD (SD36/EM) DNA fragment.
XX      DE
XX      Encapsidation; transfer vector; nephrotropic; antiparkinsonian; anti-HIV;
XX      KW      cytosatic; gene therapy; transgenic; retroviral packaging;
XX      KW      gene delivery; Parkinson's disease; infectious diseases; cancer; ds.
XX      XX
XX      Synthetic.
XX      OS
XX      WO2000040741-A2.
XX      PN
XX      13-JUL-2000.
XX      PD
XX      06-JAN-2000; 2000WO-US000390.
XX      PF
XX      07-JAN-1999; 99US-0115247P.
XX      PR
XX      (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX      PA
XX      Arya SK;
XX      PI
XX      WPI; 2000-475836/41.
XX      DR
XX      New lentivirus transfer vector, functionally deleted for a splice donor
XX      PT      site and comprising a packaging signal and transgene operably linked to a
XX      PT      promoter, for improving encapsidation or transgene RNA and for gene
XX      PT      therapy.
XX      XX
XX      Example 8; Page 125-130; 143pp; English.
XX      PS
XX      This invention describes a novel transfer vector derived from a
XX      CC      lentivirus, functionally deleted for a splice donor site (SD), and
XX      CC      comprising a packaging signal and transgene operably linked to a
XX      CC      promoter. The products of the invention have nephrotropic,
XX      CC      antiparkinsonian, anti-HIV, and cytostatic activity and can be used for
XX      CC      gene therapy. Encapsidation of transgene RNA is improved using the new
XX      CC      retroviral packaging and transfer vectors. The new transfer and packaging
XX      CC      vectors are used as gene delivery agents and allow transfer of a
XX      CC      transgene into the genome of non-dividing cells. They can be used to
XX      CC      create a high-efficiency packaging cell line that provides greatly
XX      CC      enhanced packaging of foreign DNA. Individuals suffering from a
XX      CC      deficiency in alpha-galactosidase expression, such as Fabry disease can
XX      CC      be treated by delivering the vectors to cells in vitro or in vivo.
XX      CC      Parkinson's disease, infectious diseases, such as acquired
XX      CC      immunodeficiency syndrome and cancers can be treated with the vectors.
XX      CC      The non-infective packaging vectors can be used to detect wild-type HIV
XX      CC      in biological samples using southern or northern blot assays. The
XX      CC      packaging of the viral RNA is maximised, without an increase in the
XX      CC      of the 5' SD region of the HIV-2 packaging vector results in suppressed
XX      CC      encapsidation of the packaging vector genomes without critical loss of
XX      CC      gene expression. Functional deletion of the SD site of the transfer
XX      CC      vector results in enhanced encapsidation of the transfer vector's genome.
XX      CC      HIV-2 packaging vector specifically and faithfully packages its own
XX      CC      optimally constructed transfer vector and gives better quality and titre
XX      CC      of vector than HIV-1
XX      XX
XX      Sequence 9663 BP; 3276 A; 1952 C; 2310 G; 2074 T; 0 U; 51 Other;
XX      SQ      Query Match      84.0%; Score 16.8; DB 3; Length 9663;
XX      Best Local Similarity 90.0%; Pred. No. 1.1e+03;
XX      Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX      QY      1 CAAGTAAATGCAGAACAGG 20
XX      Db      4818 CAAGTAAATGCAGAACACTAGG 4837
XX      RESULT 30
XX      ADL23697
XX      ID      ADL23697 standard; DNA; 9670 BP.
XX      AC
XX      ADL23697;
XX      XX

```

```
XX 20-MAY-2004 (first entry)
XX Human immunodeficiency virus 2 (HIV-2) ROD isolate genomic DNA.
DE ds; HIV-2 infection; LAV-II; lymphadenopathy associated virus II;
KW immunogenic; antigenic; anti-HIV; vaccine; gene.
XX Human immunodeficiency virus type 2 (ISOLATE ROD).
XX US2003235835-A1.
XX 25-DEC-2003.
XX 25-NOV-2002; 2002US-00302947.
XX 22-JAN-1986; 86FR-00000911.
PR 06-FEB-1986; 86FR-00001635.
PR 13-FEB-1986; 86FR-00001985.
PR 03-MAR-1986; 86US-00835228.
PR 18-MAR-1986; 86FR-00003881.
PR 24-MAR-1986; 86FR-00004215.
PR 06-OCT-1986; 86US-00916080.
PR 21-NOV-1986; 86US-00933184.
PR 16-JAN-1987; 87US-00003764.
PR 11-FEB-1987; 87US-00013477.
PR 03-SEP-1991; 91US-00752368.
PR 20-DEC-1991; 91US-00810908.
XX (INSP ) INST PASTEUR.
XX Alison M, Montagnier L, Guetard D, Clavel F, Sonigo P, Guyader M;
XX WPI; 2004-070575/07.
DR P-PSDB; ADL23698, ADL23699, ADL23700, ADL23701, ADL23702, ADL23703,
DR ADL23704, ADL23705, ADL23706, ADL23707, ADL23708.
XX Diagnosing HIV-2 infection by contacting genetic DNA from sample of a
PT person suspected of having HIV-2 infection with a DNA probe derived from
PT genome of HIV-2 virus and determining formation of hybridized complex.
XX Example 4; Page 6-14; 29pp; English.
XX This invention relates to diagnosing an HIV-2 (or LAV-II, lymphadenopathy
CC associated virus II) infection using an HIV-2 hybridisation DNA probe.
CC Specifically, it provides an in vitro kit capable of diagnosing an
CC infection caused by the HIV-2 virus that comprises contacting a
CC biological fluid with an immunogenic or antigenic peptide (or conjugate
CC thereof) and detecting an antigen-antibody complex by physical or
CC chemical methods. The present invention describes a detection step
CC performed by radioimmunoassays (RIA), radioimmuno-precipitation assays
CC (RIPA), immunofluorescence assays (IFA) or enzyme-linked immunosorbent
CC assays (ELISA). Accordingly, these anti-HIV compositions are useful for
CC developing vaccines to induce production of antibodies against infection
CC by HIV-2 viruses. This polynucleotide sequence is the HIV-2 ROD isolate
CC genomic DNA sequence of the invention.
XX
XX Sequence 9670 BP; 3310 A; 1992 C; 2382 G; 1983 T; 0 U; 3 Other;
SQ Query Match 84.0%; Score 16.8; DB 12; Length 9670;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20
Db |||||
4214 CAAGTAAATGCAGAACTAGG 4233

RESULT 31
AAQ20616
ID AAQ20616 standard; DNA; 9672 BP.
XX
XX AAQ20616;
AC
```

```
XX 09-SEP-2004 (revised)
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 19-MAY-1992 (first entry)
XX ROD HIV-2 isolate complete genome.
DE Human immunodeficiency virus; AIDS; ss.
XX Human immunodeficiency virus 2.
XX Unidentified.
XX Key Location/Qualifiers
PH LTR 1..173
FT /tag= a
FT /note= "LTR:R 1"
FT misc_feature 174..299
FT /tag= b
FT /note= "U5"
FT CDS 546..2111
FT /tag= c
FT /note= "GAG gene"
FT misc_feature 951..1640
FT /tag= d
FT /note= "GAG p 26"
FT misc_feature 1701..2111
FT /tag= e
FT /note= "GAG p 12"
FT misc_feature 1829..4936
FT /tag= f
FT /note= "polymerase"
FT misc_feature 4869..5513
FT /tag= g
FT /note= "O protein"
FT misc_feature 5344..5679
FT /tag= h
FT /note= "X protein"
FT misc_feature 5682..5996
FT /tag= i
FT /note= "Y protein"
FT exon 5845..6140
FT /tag= j
FT /note= "TAT gene exon 1"
FT exon 6071..6140
FT /tag= k
FT /note= "ART gene exon 1"
FT misc_feature 6147..8720
FT /tag= l
FT /note= "Env protein"
FT exon 8307..8536
FT /tag= m
FT /note= "ART gene exon 2"
FT misc_feature 8557..9324
FT /tag= n
FT /note= "F protein"
FT misc_feature 8942..9497
FT /tag= o
FT /note= "U3"
FT LTR 9498..9671
FT /tag= p
FT /note= "LTR:R 2"
XX
XX US079342-A.
XX 07-JAN-1992.
XX
XX 11-FEB-1987; 87US-00013477.
XX
XX 03-MAR-1986; 86US-00835228.
PR 06-OCT-1986; 86US-00916080.
XX 21-NOV-1986; 86US-00933184.
PR 16-JAN-1987; 87US-00003764.
```







```
XX AC ADC87444;
XX DT 01-JAN-2004 (first entry)
XX DE Human GPCR gene SEQ ID NO:1897.
XX KW ds; gene; human; GPCR;
XX OS guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX PN Homo sapiens.
XX PD EP1270724-A2.
XX PF 02-JAN-2003.
XX PR 18-JUN-2002; 2002EP-00013517.
XX PR 18-JUN-2001; 2001JP-00246789.
XX PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX WPI; 2003-315783/31.
XX DR P-PSDB; ADC87445.
XX PT New polynucleotide, useful for preparing a composition for treating a
XX PT patient in need of increased or suppressed activity or expression of the
XX PT guanosine triphosphate-binding protein coupled receptor.
XX PS Claim 1; SEQ ID NO 1897; 28pp; English.
XX CC The invention relates to a novel polynucleotide encoding a guanosine
XX CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
XX CC the invention may have a use in gene therapy. The polynucleotide and
XX CC polypeptide are useful for preparing a composition for treating a patient
XX CC in need of increased or suppressed activity or expression of the
XX CC guanosine triphosphate-binding protein coupled receptor. The
XX CC polynucleotide sequences shown in ADC85548-ADC87615 encode GPCR's of the
XX CC invention.
XX SQ Sequence 16270 BP; 4172 A; 3573 C; 4062 G; 4463 T; 0 U; 0 Other;
XX
XX Query Match 84.0%; Score 16.8; DB 10; Length 16270;
XX Best Local Similarity 90.0%; Pred. No. 1.1e+03;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 CAAGTAAATGCAGAAACAGG 20
XX DB 6075 CAAGTAAATGCAGAAACAGG 6056
XX
XX RESULT 37
XX ID ABD33401
XX AC ABD33401;
XX XX
XX DT 18-NOV-2004 (first entry)
XX DE Human cancer-associated (CA) gene HD07-075.
XX KW Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
XX KW ds; cancer; cytosstatic.
XX OS Homo sapiens.
XX PN WO2004058146-A2.
XX PD 15-JUL-2004.
XX XX

XX PF 15-DEC-2003; 2003WO-US040081.
XX PR 17-DEC-2002; 2002US-00322281.
XX PA (SAGR-) SAGRES DISCOVERY INC.
XX PI Morris DW, Malandro MS;
XX WPI; 2004-499109/47.
XX DR Novel human cancer associated protein encoded within open reading frame
XX PT of cancer associated gene, useful as targets for diagnosing cancer.
XX PS Claim 16; SEQ ID NO 516; 182pp; English.
XX CC The invention relates to cancer-associated proteins (CAP) and the cancer-
XX CC associated (CA) nucleic acids encoding them. The invention also relates
XX CC to a method for treating cancers involving administering to a patient an
XX CC inhibitor of CAP, and a method of screening for anticancer activity in a
XX CC potential drug involving providing a cell that expresses a CA gene,
XX CC contacting a tissue sample derived from a cancer cell with an anticancer
XX CC drug candidate and monitoring the effect of the anticancer drug candidate
XX CC on expression of the CA gene. The CAP proteins are useful for detecting
XX CC cancer associated with expression of a CAP protein in a test cell sample
XX CC and for screening for a bioactive agent capable of modulating the
XX CC activity of a CAP protein. The CA nucleic acids are useful for diagnosing
XX CC cancer, involving determining the expression of a CA nucleic acid in a
XX CC tissue. This sequence represents a human CA gene of the invention. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 41079 BP; 8798 A; 10965 C; 11900 G; 9118 T; 0 U; 298 Other;
XX
XX Query Match 84.0%; Score 16.8; DB 13; Length 41079;
XX Best Local Similarity 90.0%; Pred. No. 1.2e+03;
XX Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 CAAGTAAATGCAGAAACAGG 20
XX DB 2956 CATGTAAATGCAGAAACATG 2975
XX
XX RESULT 38
XX ID ADZ13504
XX AC ADZ13504;
XX DT 16-JUN-2005 (first entry)
XX DE Human cancer-associated genomic DNA #87.
XX KW Diagnosis; DNA microarray; microarray; biochip; cancer; neoplasm;
XX KW cytosstatic; gene; ds.
XX OS Homo sapiens.
XX PN WO2005031001-A2.
XX PD 07-APR-2005.
XX PF 23-SEP-2004; 2004WO-US031617.
XX PR 23-SEP-2003; 2003US-00669920.
XX PA (CHIR ) CHIRON CORP.
XX PI Morris DW, Malandro MS;
XX WPI; 2005-273395/28.
XX PT Nucleic acid array useful for detecting cancer associated nucleic acid,
```





XX 27-MAR-2001; 2001GB-00007658.  
 XX (CHIR-) CHIRON SPA.  
 XX (GENO-) INST GENOMIC RES.  
 XX Maignani V, Tettelin H, Fraser C;  
 XX WPI; 2003-040579/03.  
 XX P-PSDB; ABU02823.  
 XX  
 XX New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,  
 XX useful as medicaments for treating or preventing a disease or infection  
 XX due to *Streptococcus* bacteria, such as pneumonia, sepsis, otitis media or  
 XX ear infection.  
 XX  
 XX Claim 6; SEQ ID NO 4807; 56pp; English.  
 XX  
 XX The invention relates to a protein comprising or having at least 50%  
 XX identity to any of the 2469 amino acid sequences, identified in the  
 XX specification (available on a computer readable format), or its fragment,  
 XX expressed from 2469 of 2489 identified DNA coding regions from the  
 XX *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as  
 XX AB556454. Also included are an antibody which binds one of the proteins,  
 XX treating a patient by administering the protein, DNA or antibody (in a  
 XX composition), a kit comprising first and second primers, which are the  
 XX nucleic acid cited above or fragments between nucleotides 8-100 of a  
 XX sequence not defined in the specification, for amplifying a target  
 XX sequence contained within a *Streptococcus* nucleic acid sequence, where  
 XX the first primer is substantially complementary to the target sequence  
 XX and the second primer is substantially complementary to the complement of  
 XX the target sequence, and where the parts of the primers having  
 XX substantial complementarity define the termini of the target sequence to  
 XX be amplified, assay comprising contacting a test compound with the  
 XX protein, and determining whether the test compound binds to the protein  
 XX and a *Streptococcus pneumoniae* bacterium, where one or more genes  
 XX encoding the proteins has been rendered inactive. The proteins, nucleic  
 XX acid molecules, antibody and compositions are useful as medicaments for  
 XX treating or preventing a disease or infection due to *Streptococcus*  
 XX bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis  
 XX media or ear infection. They are also useful in developing vaccines,  
 XX diagnostics and antibiotics. The methods are useful for identifying  
 XX immunodominant proteins. The present sequence is one of the 2489  
 XX identified coding regions from the genomic sequence. Note: The sequence  
 XX data for this patent did not form part of the printed specification, but  
 XX was obtained in electronic format directly from WIPO at  
 XX ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
 XX standardise OS field)  
 XX  
 XX Sequence 507 BP; 149 A; 64 C; 85 G; 209 T; 0 U; 0 Other;  
 Query Match 82.0%; Score 16.4; DB 10; Length 507;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AAGTAAATGCAGAAACAG 19  
 |||||  
 DB 501 AAGTAAATGCAGTAAACAG 484  
 RESULT 43  
 ADM91806/c  
 ID ADM91806 standard; DNA; 507 BP.  
 XX  
 XX ADM91806;  
 XX  
 XX 03-JUN-2004 (first entry)  
 XX  
 XX S pneumoniae antigenic protein-encoding gene sequence SeqID3.  
 XX antibacterial; gene therapy; *Streptococcus pneumoniae* infection;  
 XX antigenic; gene; ds.  
 XX

OS *Streptococcus pneumoniae*.  
 XX WO2004020609-A2.  
 XX  
 XX 11-MAR-2004.  
 XX  
 XX 02-SEP-2003; 2003WO-US027401.  
 XX  
 XX 30-AUG-2002; 2002US-0407082P.  
 XX  
 XX (TUFT ) UNIV TUFTS.  
 XX  
 XX Camilli A, Hava DL;  
 XX WPI; 2004-239189/22.  
 XX P-PSDB; ADM92043.  
 XX  
 XX New *Streptococcus pneumoniae* nucleic acid molecules, useful for  
 XX diagnosing, treating and preventing active infections of *Streptococcus*  
 XX pneumoniae.  
 XX  
 XX Claim 1; SEQ ID NO 3; 123pp; English.  
 XX  
 XX This invention relates to novel isolated *Streptococcus pneumoniae* nucleic  
 XX acid molecules and the antigenic polypeptides encoded by them. The  
 XX invention may be useful for the production of compounds with an  
 XX antibacterial activity or for gene therapy. The nucleic acid molecules,  
 XX compositions and methods disclosed are useful for treating *Streptococcus*  
 XX pneumoniae infection. The present sequence is that of an S pneumoniae  
 XX gene of the invention.  
 XX  
 XX Sequence 507 BP; 149 A; 64 C; 85 G; 209 T; 0 U; 0 Other;  
 Query Match 82.0%; Score 16.4; DB 12; Length 507;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AAGTAAATGCAGAAACAG 19  
 |||||  
 DB 501 AAGTAAATGCAGTAAACAG 484  
 RESULT 44  
 ACH75526  
 ID ACH75526 standard; DNA; 512 BP.  
 XX  
 XX ACH75526;  
 XX  
 XX 29-JUL-2004 (first entry)  
 XX  
 XX Human genome derived single exon probe #8721.  
 XX  
 XX Human; probe; ss; gene expression; single exon probe; microarray;  
 XX alternative splicing event; genomic alteration.  
 XX  
 XX Homo sapiens.  
 XX  
 XX US2003194704-A1.  
 XX  
 XX 16-OCT-2003.  
 XX  
 XX 03-APR-2002; 2002US-00029386.  
 XX  
 XX 03-APR-2002; 2002US-00029386.  
 XX  
 XX (PENN/) PENN S G.  
 XX (RANK/) RANK D R.  
 XX (HANZ/) HANZEL D K.  
 XX  
 XX Penn SG, Rank DR, Hanzel DK;  
 XX WPI; 2004-119264/12.  
 XX

PT New human genome-derived single exon nucleic acid probes useful for human  
PT gene expression analysis, for identifying or characterizing alternative  
PT splicing events, for assessing genomic alterations or as tools for  
PT surveying tissues.

XX Claim 15; SEQ ID NO 8721; 80pp; English.

XX The invention relates to a nucleic acid probe for measuring human gene  
CC expression, comprising any of the 27,400 fully defined nucleotide  
CC sequences in the specification, or their complements or fragments, and  
CC encoding at least 8 amino acids of any of the 688 amino acid sequences  
CC fully defined in the specification. The probe is a single exon probe that  
CC hybridizes under high stringency conditions to a nucleic acid molecule  
CC expressed in human cells or tissues. Also included are a spatially-  
CC addressable set of single exon nucleic acid probes for measuring human  
CC gene expression (comprising a plurality of single exon nucleic acid  
CC probes cited above, where each of the plurality of probes is separately  
CC and addressably isolatable or amplifiable from the plurality), a single  
CC exon microarray for measuring human gene expression, a method of  
CC measuring human gene expression, a vector comprising the single exon  
CC probe cited above, an ORF-encoded peptide comprising at least 8  
CC contiguous amino acids of any of the above-mentioned amino acid  
CC sequences (optionally with conservative amino acid substitutions), an  
CC isolated antibody that binds specifically to a peptide cited above,  
CC methods of selling and/or licensing single exon probes or microarrays to  
CC a customer desiring to measure gene expression, a method of providing  
CC human gene expression data by subsequence, and a computer-readable  
CC storage medium which contains a database having a plurality of records  
CC (each record including data on the expression of a single exon probe  
CC cited above. The probe, methods and apparatus are useful in gene  
CC expression analysis. The probes may be used as tools for surveying  
CC tissues to detect the presence of expressed messages that contain their  
CC specific exon, or in constructing genome-derived single exon microarrays.  
CC In addition, the probes are used in identifying and characterizing  
CC alternative splicing events, in detecting and characterizing gross  
CC alterations in the genomic locus that includes their exon, in assessing  
CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
CC or in expressing the ORF-encoded peptide. The present sequence is a human  
CC single exon probe of the invention. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?DocID=20030194704

XX Sequence 512 BP; 130 A; 86 C; 131 G; 165 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 12; Length 512;  
Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19

DB 331 AAGCAATGCAGAAACAG 348

RESULT 45  
ACN52635/c  
ID ACN52635 standard; cDNA; 626 BP.

XX ACN52635;

DT 02-DEC-2004 (first entry)

XX Cotton androecium tissue EST Clone ID: LIB3828-017-Q1-K6-B7, SEQ:7416.

XX Cotton; plant; EST; expressed sequence tag; transgenic plant; androecium;  
KW variety Nuotton33B; library LIB3828; molecular tag; molecular marker;  
KW genetic mapping; molecular mapping; seed germination; plant growth;  
KW plant quality; plant yield; plant breeding; tissue printing; ss.

OS Gossypium hirsutum.

XX US2004123340-A1.

XX

PD 24-JUN-2004.

XX 12-DEC-2001; 2001US-00021323.

XX 14-DEC-2000; 2000US-0255619P.

XX (DEIK/) DEIKMAN J.

PA (FENG/) FENG P C C.

PA (FINC/) FINCHER K L.

XX (ZIEG/) ZIEGLER T E.

PI Deikman J, Feng PCC, Fincher KL, Ziegler TE;

XX WPI; 2004-479808/45.

XX New isolated nucleic acid molecule that encodes a plant protein or its  
PT fragment, useful for isolating a variety of agronomically significant  
PT genes associated with plant growth, quality or yield, and as molecular  
PT tags to map genes.

XX Claim 1; SEQ ID NO 7416; 34pp; English.

XX The invention relates to 17880 cotton expressed sequence tags (ESTs;  
CC ACN45220-ACN63099). The ESTs were isolated from cDNA libraries generated  
CC from primed or non-primed seeds from variety DP50B, mature seeds from  
CC variety Coker 312 Beswell 96 Field, and androecium tissue, gynoecium  
CC tissue, developing fibres, carpel walls and septa from variety  
CC Nuotton33B. The invention also relates to substantially purified  
CC proteins or their fragments encoded by nucleic acid molecules of the  
CC invention, and to transformed plants having a nucleic acid construct  
CC comprising a nucleic acid of the invention. The cotton ESTs are useful as  
CC molecular tags to isolate genetic regions, to isolate genes, to map  
CC genes, to determine gene function and to determining whether genes are  
CC members of a particular gene family. The nucleic acid molecules may be  
CC used for isolating a variety of agronomically significant genes  
CC associated with plant growth, quality, yield, and could also serve as  
CC links in metabolic and catabolic pathways. The nucleic acid molecules are  
CC also useful for identifying genes important in initiating and maintaining  
CC seed germination or that may be used to mitigate stresses encountered  
CC during seed germination. The ESTs additionally enable the acquisition of  
CC promoters and cis-regulatory elements which will be useful to express  
CC agronomically significant genes in these tissues and/or other tissues,  
CC and also permits the acquisition of molecular markers useful in breeding  
CC schemes, genetic and molecular mapping, and in cloning of agronomically  
CC significant genes. The nucleic acid molecules are further useful for  
CC detecting the expression level or pattern of a protein or mRNA and for  
CC detecting the presence or quantity of a protein by tissue printing. The  
CC present sequence represents a specifically claimed EST isolated from a  
CC cotton variety Nuotton33B androecium tissue cDNA library (LIB3828). The  
CC cotton variety Nuotton33B androecium tissue cDNA library (LIB3828). The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the US  
CC patent office at seqdata.uspto.gov/sequence.html?DocID=US20040123340

XX Sequence 626 BP; 194 A; 89 C; 101 G; 240 T; 0 U; 2 Other;

Query Match 82.0%; Score 16.4; DB 13; Length 626;  
Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACA 18

DB 274 CAAGTAAAGCAGAAACA 257

RESULT 46

ACF73876/c

ID ACF73876 standard; DNA; 711 BP.

XX ACF73876;

XX 20-NOV-2003 (first entry)

XX Staphylococcus aureus DNA #1556.

XX Antibacterial; vaccine; gene therapy; infection; sepsis; diagnosis;  
 KW enzymatic assay; antibiotic target; gene; ds.  
 XX Staphylococcus aureus.  
 OS WO200294868-A2.  
 PN 28-NOV-2002.  
 PD 27-MAR-2002; 2002WO-IB002637.  
 PF 27-MAR-2001; 2001GB-00007661.  
 XX (CHIR-) CHIRON SPA.  
 XX Masignani V, Mora M, Scarselli M;  
 PI WPI; 2003-120786/11.  
 DR P-PSDB; ABM72316.  
 XX New Staphylococcus aureus protein, useful as a vaccine for treating or  
 PT preventing Staphylococcal infection, specifically an infection caused by  
 PT S. aureus, e.g. sepsis.  
 XX Claim 6; SEQ ID NO 3111; 49pp; English.  
 PS The invention relates to novel genes and encoded proteins from  
 CC Staphylococcus aureus. A composition comprising the S. aureus protein, a  
 CC nucleic acid encoding the protein, or an antibody to the protein, is  
 CC useful as a pharmaceutical, particularly as a vaccine for treating or  
 CC preventing infection due to Staphylococcus bacteria, specifically an  
 CC infection caused by S. aureus. The composition is particularly useful for  
 CC treating or preventing sepsis in a patient. The composition can also be  
 CC used for diagnostics. The protein is also used in an assay for enzymatic  
 CC studies and as a target for antibiotics. This sequence represents one of  
 CC the novel S. aureus genes of the invention  
 XX  
 SQ Sequence 711 BP; 241 A; 117 C; 141 G; 212 T; 0 U; 0 Other;  
 Query Match 82.0%; Score 16.4; DB 8; Length 711;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACA 18  
 |||||  
 Db 137 CAAGTAAATGCAGAAATCA 120  
 RESULT 47  
 ABN93331/c  
 ID ABN93331 standard; DNA; 714 BP.  
 AC ABN93331;  
 XX 24-JUL-2002 (first entry)  
 DT Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2794.  
 DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 XX antibacterial; gene therapy; gene; ds.  
 KW Staphylococcus epidermidis.  
 XX US6380370-B1.  
 OS 30-APR-2002.  
 PN 13-AUG-1998; 98US-00134001.  
 PF 14-AUG-1997; 97US-0055779P.  
 XX 08-NOV-1997; 97US-0064964P.  
 PR

PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX Doucette-Stamm LA, Bush D;  
 PI WPI; 2002-381255/41.  
 DR P-PSDB; ABP40786.  
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermis  
 PT polypeptide, useful for diagnosing and treating bacterial infections.  
 XX Disclosure; SEQ ID NO 2794; 267pp; English.  
 PS ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences can  
 CC also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life cycle  
 CC or inhibit S. epidermidis infection. N.B. The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from the USPTO web site  
 XX Sequence 714 BP; 247 A; 108 C; 137 G; 222 T; 0 U; 0 Other;  
 SQ Query Match 82.0%; Score 16.4; DB 6; Length 714;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACA 18  
 |||||  
 Db 137 CAAGTAAATGCAGAAATCA 120  
 RESULT 48  
 ADS04339/c  
 ID ADS04339 standard; DNA; 714 BP.  
 XX ADS04339;  
 AC 04-NOV-2004 (first entry)  
 DT Staphylococcus epidermis polynucleotide seqid 3634.  
 DE antibacterial; vaccine; antisense therapy; Staphylococcus epidermidis;  
 KW recombinant expression vector; infection; computer readable medium;  
 KW computer based system; gene; ds.  
 XX Staphylococcus epidermidis.  
 OS US2004147734-A1.  
 PN 29-JUL-2004.  
 PD 01-DEC-2003; 2003US-00724972.  
 PF 08-NOV-1997; 97US-0064964P.  
 XX 13-AUG-1998; 98US-00134001.  
 PR 29-NOV-1999; 99US-00450969.  
 XX (DOUC/) DOUCETTE-STAMM L.  
 PA (BUSH/) BUSH D.  
 XX Doucette-Stamm L, Bush D;  
 PI WPI; 2004-580138/56.  
 DR P-PSDB; ADS08111.  
 XX New isolated polypeptide and encoding nucleic acid derived from  
 PT Staphylococcus epidermidis, useful for diagnosing, preventing and/or  
 PT treating an S. epidermidis bacterial infection.  
 XX Claim 5; SEQ ID NO 3634; 741pp; English.  
 PS



XX The invention describes an isolated nucleic acid comprising a nucleotide  
 CC sequence with any of 3772 fully defined nucleotide sequences (SEQ ID NO:  
 CC 1-3772) and encoding an *Staphylococcus epidermidis* polypeptide with any  
 CC of 3772 fully defined amino acid sequences (SEQ ID NO: 3772-7544) as  
 CC given in the specification. Also described are: a recombinant expression  
 CC vector; a cell comprising a recombinant expression vector of (1);  
 CC producing an *S. epidermidis* polypeptide; an isolated nucleic acid  
 CC comprising a nucleotide sequence of at least 8 nucleotides in length; a  
 CC vaccine composition for prevention or treatment of an *S. epidermidis*  
 CC infection, comprising a nucleic acid cited above and a carrier; treating  
 CC a subject for *S. epidermidis* infection; a recombinant or substantially  
 CC pure preparation of an *S. epidermidis* polypeptide or its fragment; a  
 CC vaccine composition for prevention or treatment of an *S. epidermidis*  
 CC infection; detecting the presence of a *Staphylococcus* nucleic acid in a  
 CC sample; a computer readable medium having recorded in it the nucleotide  
 CC sequences with SEQ ID NO: 1-3772 or its fragments; a computer based  
 CC system for identifying fragments of the *Staphylococcus* genome of  
 CC commercial importance; a computer based system for identifying fragments  
 CC of the *Staphylococcus* plasmids of commercial importance; identifying  
 CC commercially important nucleic acid fragments of the *Staphylococcus*  
 CC genome and/or plasmids; and identifying an expression modulating fragment  
 CC of the *Staphylococcus* genome and/or plasmids. The methods and  
 CC compositions of the present invention are useful for the diagnosis,  
 CC prevention and/or treatment of an *Staphylococcus epidermidis* bacterial  
 CC infection. This sequence encodes a *S. epidermidis* protein of the invention.

XX Sequence 714 BP; 247 A; 108 C; 137 G; 222 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 13; Length 714;

Best Local Similarity 94.4%; Pred. No. 1.3e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCGAAGACA 18

Db 137 CAAGTAATGCGAAGATCA 120

RESULT 49

ACA24380

ID ACA24380 standard; DNA; 762 BP.

XX ACA24380;

XX 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #6037.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.

XX Bacteroides fragilis.

XX WO200277183-A2.

XX 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

XX 25-OCT-2001; 2001US-0342923P.

XX 08-FEB-2002; 2002US-00072851.

XX 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

XX P-PSDB; ABU20510.

PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 12250; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC the gene product or that has an activity against a biological pathway  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 762 BP; 274 A; 147 C; 180 G; 161 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 8; Length 762;

Best Local Similarity 94.4%; Pred. No. 1.3e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCGAAGACA 19

Db 296 AAGTAATGCGAAGACA 313

RESULT 50

AAV75056

ID AAV75056 standard; DNA; 924 BP.

XX AAV75056;

XX 16-MAR-1999 (first entry)

XX Staphylococcus aureus contig SEQ ID #745.

XX Computer readable medium; vaccine; *S. aureus* infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.

XX Staphylococcus aureus.

XX Key Location/Qualifiers

FT misc\_feature 481..540

FT /\*tag= a

FT /note= "these bases represent a line of missing text in  
 FT the sequence listing in the specification. They are  
 FT included to maintain the nucleotide numbering given in  
 FT the specification for this DNA sequence"



XX EP786519-A2.  
 PN 30-JUL-1997.  
 PD 07-JAN-1997; 97EP-00100117.  
 XX 05-JAN-1996; 96US-0009861P.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX Kunsch CA, Choi GH, Baraah SC, Dillon PJ, Fannon MR, Rosen CA;  
 PI WPI; 1997-374922/35.  
 XX Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -  
 PT stored on computer readable medium and used in the production of anti-  
 PT *S.aureus* vaccines.  
 XX Claim 1; Page 1641-1642; 3271pp; English.  
 XX This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the *S.aureus* DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against *S.aureus* infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the *S.aureus* DNA sequences contained on the computer  
 CC readable medium  
 XX  
 XX Sequence 924 BP; 258 A; 168 C; 148 G; 289 T; 0 U; 61 Other;  
 SQ

Query Match 82.0%; Score 16.4; DB 2; Length 924;  
 Best Local Similarity 94.4%; Pred. No. 1.3e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAAGTAAATGCAGAAACA 18  
 |||||  
 DB 273 CAAAGTAAATGCAGAAACA 290  
 |||||

RESULT 51  
 AAT43303  
 ID AAT43303 standard; DNA; 1035 BP.  
 XX AAT43303;  
 AC  
 XX 10-FEB-1997 (first entry)  
 DT  
 XX Vmp7 soluble variant coding sequence.  
 DE  
 XX OspA; OspB; outer surface protein A; Borrelia; variable major protein 7;  
 KW Borrelia hermslii; Vmp7; surface lipoprotein; spirochete; human; antigen;  
 KW Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;  
 KW arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;  
 KW ds.  
 XX Borrelia hermslii.  
 OS  
 XX US5571718-A.  
 PN  
 XX 05-NOV-1996.  
 PD  
 XX 08-SEP-1992; 92US-00941523.  
 PF  
 XX 21-DEC-1990; 90US-00632072.  
 PR  
 XX (ASUY-) ASSOC UNIVERSITIES INC.  
 PA  
 XX Barbour AG, Dunn JJ;  
 PI

PF 08-SEP-1992; 92US-00941523.  
 XX 21-DEC-1990; 90US-00632072.  
 PR (ASUY-) ASSOC UNIVERSITIES INC.  
 XX Barbour AG, Dunn JJ;  
 PI WPI; 1996-505409/50.  
 XX P-PSDB; AAW08097.  
 XX Soluble recombinant forms of Borrelia lipo:proteins - useful for vaccine  
 PT prodn. for treatment of Lyme disease.  
 PT Claim 4; Col 43-46; 49pp; English.  
 XX AAT43301-743303 represent coding sequences for soluble recombinant  
 CC Borrelia proteins of the invention. This sequence represents the coding  
 CC sequence for the soluble recombinant variant of the Borrelia hermslii  
 CC variable major protein 7 (Vmp7). Borrelia spirochetes are responsible for  
 CC a variety of human disorders including Lyme borreliosis, and relapsing  
 CC fevers. The spirochete is transmitted to humans and animals through the  
 CC bite of a tick, and can cause serious dermatological, arthritic,  
 CC neurological and other pathogenic disorders in an infected host. This  
 CC sequence is used to create recombinant host cells, and the encoded Vmp7  
 CC protein can be isolated from the cytosol of one of these cells without  
 CC the use of detergent. The encoded recombinant proteins can be used as  
 CC antigens for the production of vaccines against Lyme disease. The  
 CC recombinant proteins can also be used in immunoassays and other  
 CC diagnostic screening methods to detect the presence of antibodies against  
 CC Borrelia lipoproteins in the sera of infected patients  
 XX  
 XX Sequence 1035 BP; 397 A; 141 C; 274 G; 223 T; 0 U; 0 Other;  
 SQ

Query Match 82.0%; Score 16.4; DB 2; Length 1035;  
 Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAGG 20  
 |||||  
 DB 246 AGTAAATGCAGAAACAGG 263  
 |||||

RESULT 52  
 AAT43316  
 ID AAT43316 standard; DNA; 1110 BP.  
 XX AAT43316;  
 AC  
 XX 10-FEB-1997 (first entry)  
 DT  
 XX Vmp7 variant #1 coding sequence.  
 DE  
 XX OspA; OspB; outer surface protein A; Borrelia; variable major protein 7;  
 KW Borrelia hermslii; Vmp7; surface lipoprotein; spirochete; human; antigen;  
 KW Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;  
 KW arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;  
 KW ds.  
 XX Synthetic.  
 OS  
 XX US5571718-A.  
 PN  
 XX 05-NOV-1996.  
 PD  
 XX 08-SEP-1992; 92US-00941523.  
 PF  
 XX 21-DEC-1990; 90US-00632072.  
 PR  
 XX (ASUY-) ASSOC UNIVERSITIES INC.  
 PA  
 XX Barbour AG, Dunn JJ;  
 PI

```

DR WPI; 1996-505409/50.
XX Soluble recombinant forms of Borrelia lipo:proteins - useful for vaccine
PT prodn. for treatment of Lyme disease.
XX Example 9; Col 41-42; 49pp; English.
XX This sequence represents the coding sequence for a soluble recombinant
CC Borrelia variable major protein 7 (Osp7) variant protein of the
CC invention. Borrelia spirochetes are responsible for a variety of human
CC disorders including Lyme borreliosis, and relapsing fevers. The
CC spirochete is transmitted to humans and animals through the bite of a
CC tick, and can cause serious dermatological, arthritic, neurological and
CC other pathogenic disorders in an infected host. This sequence is used to
CC create recombinant host cells, and the encoded vmp7 protein can be
CC isolated from the cytosol of one of these cells without the use of
CC detergent. The encoded recombinant proteins can be used as antigens for
CC the production of vaccines against Lyme disease. The recombinant proteins
CC can also be used in immunoassays and other diagnostic screening methods
CC to detect the presence of antibodies against Borrelia lipoproteins in the
CC sera of infected patients
XX
SQ Sequence 1110 BP; 432 A; 143 C; 286 G; 249 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 2; Length 1110;
Best Local Similarity 94.4%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAGG 20
DB 321 AGTAAACGACGAAACAGG 338

RESULT 53
AAA26350/c
ID AAA26350 standard; cDNA; 1196 BP.
XX
AC AAA26350;
XX
DT 29-JUN-2000 (first entry)
XX
DE Human secreted protein gene 5 SEQ ID NO:15.
XX
KW Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW antiHIV; antinflammatory; nontropic; neuroprotective; antiallergic;
KW osteopathic; antiarthritic; antibacterial; antidiabetic; antiasthma;
KW antipsoriatic; cardiant; gene therapy; cancer; neurological disorder;
KW immune disease; inflammation; blood disorder; tumour; ss.
XX
OS Homo sapiens.
XX
FN WO200006698-A1.
XX
PD 10-FEB-2000.
XX
PF 29-JUL-1999; 99WO-US017130.
XX
PR 30-JUL-1998; 98US-0094657P.
PR 05-AUG-1998; 98US-0095486P.
PR 06-AUG-1998; 98US-0095454P.
PR 06-AUG-1998; 98US-0095453P.
PR 12-AUG-1998; 98US-0096319P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Komatsoulis GA, Rosen CA, Ruben SM, Duan R, Moore PA, Shi Y;
PI Lafleur D, Wei Y, Ni J, Florence KA, Young PE, Brewer LA;
PI Soppet DR, Endress GA, Ebner R, Olsen HS, Mucenski M;
XX
DR WPI; 2000-195282/17.
DR P-PSDB; AAY91455.
XX
PT New isolated human genes and the secreted polypeptides they encode,

```

```

PT useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders.
XX Claim 1; Page 374; 634pp; English.
XX
CC The polynucleotide sequences given in AAA26346 to AAA26458 encode the
CC human secreted proteins given in AAY91451 to AAY91691. The human secreted
CC proteins can have activities based on the tissues and cells they are
CC expressed in. Examples of the activities are: cytostatic;
CC immunosuppressive; antiHIV; antinflammatory; nontropic; neuroprotective;
CC antiallergic; osteopathic; antiarthritic; antibacterial; antidiabetic;
CC antiasthma; antipsoriatic; and cardiant. The polynucleotides and their
CC corresponding secreted proteins are useful for preventing, treating or
CC ameliorating medical conditions, e.g. by protein or gene therapy. Also
CC pathological conditions can be diagnosed by determining the amount of the
CC proteins in a sample or by determining the presence of mutations in the
CC polynucleotides. Specific uses are described for each of the
CC polynucleotides, based on which tissues they are most highly expressed
CC in, and include developing products for the diagnosis or treatment of
CC cancer, tumours, neurodegenerative disorders, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, hepatic and renal disease, inflammation, allergies,
CC Alzheimer's and behavioural disorders, schizophrenia, osteoporosis,
CC arthritis, infections, AIDS, spinal cord injuries, transplant rejection,
CC diabetes, asthma, sepsis, acne, psoriasis, cardiovascular disorders,
CC reproductive disorders, gastrointestinal disorders, respiratory disorders
CC and metabolic disorders. The proteins or polynucleotides can also be used
CC as food additives or preservatives. The proteins are also useful for
CC identifying their binding partners. AAA26337 to AAA26345 and AAY91450 are
CC sequences used in the exemplification of the present invention
XX
SQ Sequence 1196 BP; 333 A; 171 C; 192 G; 500 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 3; Length 1196;
Best Local Similarity 94.4%; Pred. No. 1.4e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCAGAAACA 18
DB 311 CAATTAATGCAGAAACA 294

RESULT 54
ADL71411/c
ID ADL71411 standard; cDNA; 1196 BP.
XX
AC ADL71411;
XX
DT 20-MAY-2004 (first entry)
XX
DE Novel human secreted protein cDNA seqid 15.
XX
KW antinflammatory; neuroprotective; nontropic; antiparkinsonian;
KW anticonvulsant; antilipaeamic; CNS; gynaecological; antiarthritic;
KW antiasthmatic; anti-HIV; virucide; endocrine; cytostatic;
KW immunosuppressive; antiallergic; cardiovascular; respiratory;
KW dermatological; antimicrobial; gastrointestinal; gene therapy;
KW neurodegenerative disease; behavioural disorder; inflammatory condition;
KW hyperproliferative disorder; Alzheimer's disease; Parkinson's disease;
KW Huntington's disease; metabolic disorder; Tay-Sach's disease;
KW Leash-Nyhan syndrome; reproductive disorder; immunological disorder;
KW arthritis; asthma; AIDS; endocrine disorder; immune disorder;
KW Hodgkin's lymphoma; haematopoietic disorder; muscular disorder;
KW leukaemia; autoimmune disorder; allergy; cancer; cardiovascular disorder;
KW respiratory disorder; pulmonary disorder; connective tissue disorder;
KW skin disorder; CNS disorder; congenital disorder; infectious disorder;
KW gastrointestinal disorder; human; secreted protein; gene; ss.
XX
OS Homo sapiens.
XX
FN US2004034196-A1.
XX
PD 19-FEB-2004.

```



CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents cDNA encoding a human ovarian antigen of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1808 BP; 430 A; 464 C; 484 G; 427 T; 0 U; 3 Other;

Query Match 82.0%; Score 16.4; DB 6; Length 1808;  
 Best Local Similarity 85.0%; Pred. No. 1.4e+03;  
 Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAACTAATGCGAGAAACAG 20  
 DB 1770 CATGTAATGCGAGAAACATG 1789

RESULT 57  
 ABA90791  
 ID ABA90791 standard; DNA; 2014 BP.  
 XX ABA90791;  
 XX 14-FEB-2002 (first entry)  
 XX Bacillus anthracis DNA fingerprinting PCR primer SEQ ID NO 22.  
 KW Bacillus anthracis; MLVA primer; VNTR; DNA finger printing; anthrax;  
 KW multilocus variable number tandem repeat analysis primer; PCR primer; ss.  
 OS Bacillus anthracis.  
 XX WO200181543-A2.  
 XX 01-NOV-2001.  
 XX 26-APR-2001; 2001WO-US013373.  
 XX 26-APR-2000; 2000US-0199911P.  
 XX (REGC ) UNIV CALIFORNIA.  
 XX Keim PS, Jackson PJ;  
 XX WPI; 2002-055350/07.  
 XX Novel nucleic acid molecules, primers for multilocus variable number  
 XX tandem repeat analysis which are specific for Bacillus anthracis  
 XX templates, useful for DNA finger printing of B. anthracis.  
 XX Claim 43; Page 30; 71pp; English.  
 PS The invention relates to an isolated nucleic acid molecule (ABA90786-  
 CC ABA90790 and ABA90848-ABA90875) and a MLVA primer (multilocus VNTR  
 CC (variable number tandem repeat) analysis primer, ABA90770-ABA90785 and  
 CC ABA90791-ABA90847) specific for Bacillus anthracis templates used in DNA  
 CC finger printing of B. anthracis, useful for identifying strains of B.  
 CC anthracis and related bacteria  
 XX  
 SQ Sequence 2014 BP; 884 A; 278 C; 524 G; 328 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 6; Length 2014;  
 Best Local Similarity 94.4%; Pred. No. 1.4e+03;  
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAATGCGAGAAACAG 19  
 DB 419 AAGTAATGCGAGAAACAG 436

RESULT 58  
 ABA90848  
 ID ABA90848 standard; DNA; 2019 BP.

334 AAGTAGATGCGAGAAACAG 351  
 DB ABO54797 standard; cDNA; 1808 BP.  
 AC ABO54797;  
 XX 22-AUG-2002 (first entry)  
 DE Human ovarian antigen HTHD776 cDNA, SEQ ID NO:677.  
 XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive; gene; ss.  
 XX Homo sapiens.  
 OS  
 XX WO200200677-A1.  
 XX 03-JAN-2002.  
 XX 07-JUN-2001; 2001WO-US018569.  
 XX 07-JUN-2000; 2000US-0209467P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Birse CE, Rosen CA;  
 XX WPI; 2002-147878/19.  
 XX P-PSDB; ABP41720.  
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 XX useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
 XX cancer), immune disorders, cardiovascular disorders and neurological  
 XX diseases.  
 XX Claim 1; SEQ ID NO 677; 2922pp; English.  
 PS The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovary and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders  
 CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies

```

XX ABA90848;
AC
XX 14-FEB-2002 (first entry)
DT
XX
XX Bacillus anthracis variable region DNA sequence SEQ ID NO 79.
DE
XX
XX Bacillus anthracis; MLVA primer; VNTR; DNA finger printing; anthrax;
KW multilocus variable number tandem repeat analysis primer; ds.
XX
XX Bacillus anthracis.
OS
XX WO200181543-A2.
PN
XX 01-NOV-2001.
PD
XX
XX 26-APR-2001; 2001WO-US013373.
PF
XX
XX 26-APR-2000; 2000US-0199911P.
PR
XX (REGC ) UNIV CALIFORNIA.
PA
XX Keim PS, Jackson PJ;
PI
XX WPI; 2002-055350/07.
DR
XX
XX Novel nucleic acid molecules, primers for multilocus variable number
PT tandem repeat analysis which are specific for Bacillus anthracis
PT templates, useful for DNA finger printing of B. anthracis.
PT
XX
XX Claim 47; Page 33-34; 71pp; English.
PS
XX
XX The invention relates to an isolated nucleic acid molecule (ABA90786-
CC ABA90790 and ABA90848-ABA90875) and a MLVA primer (multilocus VNTR
CC (variable number tandem repeat) analysis primer, ABA90770-ABA90785 and
CC ABA90791-ABA90847) specific for Bacillus anthracis templates used in DNA
CC finger printing of B. anthracis, useful for identifying strains of B.
CC anthracis and related bacteria
CC
XX
XX Sequence 2019 BP; 884 A; 279 C; 524 G; 328 T; 0 U; 4 Other;
SQ
    Query Match      82.0%; Score 16.4; DB 6; Length 2019;
    Best Local Similarity 94.4%; Pred. No. 1.4e+03;
    Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 AAGTAAATGCAGAAACAG 19
Db      ||||| ||||| |||||
        419 AAGTAAATGCAGAAACAG 436

RESULT 59
ABQ67188/c
ID ABQ67188 standard; DNA; 3807 BP.
AC
XX ABQ67188;
XX
XX 29-AUG-2002 (first entry)
DT
XX
XX Listeria innocua contig DNA sequence #1.
DE
XX
XX Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
XX Listeria innocua.
OS
XX WO200228891-A2.
PN
XX
XX 11-APR-2002.
PD
XX
XX 04-OCT-2001; 2001WO-FR003061.
PF
XX
XX 04-OCT-2000; 2000FR-00012697.
PR
XX
XX
XX

```

---

```

PA (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Kunet F, Glaser P;
PI
XX WPI; 2002-332479/37.
DR
XX
XX New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators.
PT
XX
XX Claim 5; SEQ ID NO 1; 180pp; French.
PS
XX
XX The present invention relates to nucleic acid sequences (ABQ67188-
CC ABQ71212) from Listeria sp. The sequences are useful as probes and
CC primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of gene
CC expression. Proteins encoded by the nucleic acid sequences can be used to
CC screen for compounds that modulate gene expression, replication and
CC pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in anti-
CC Listeria vaccines. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3807 BP; 1392 A; 600 C; 641 G; 1174 T; 0 U; 0 Other;
SQ
    Query Match      82.0%; Score 16.4; DB 6; Length 3807;
    Best Local Similarity 94.4%; Pred. No. 1.5e+03;
    Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 AAGTAAATGCAGAAACAG 19
Db      ||||| ||||| |||||
        3464 AAGTAAATGCTGAACAG 3447

RESULT 60
ACA22058
ID ACA22058 standard; DNA; 3912 BP.
AC
XX ACA22058;
XX
XX 19-JUN-2003 (first entry)
DT
XX
XX Prokaryotic essential gene #3715.
DE
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
XX Bacillus anthracis.
OS
XX WO200277183-A2.
PN
XX
XX 03-OCT-2002.
PD
XX
XX 21-MAR-2002; 2002WO-US009107.
PF
XX
XX 21-MAR-2001; 2001US-00815242.
PR
XX 06-SEP-2001; 2001US-00948993.
PR
XX 25-OCT-2001; 2001US-0342923P.
PR
XX 08-FEB-2002; 2002US-00072851.
PR
XX 06-MAR-2002; 2002US-0362699P.
PR
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR
XX P-PSDB; ABU18188.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to

```

PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 14; SEQ ID NO 9928; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 3912 BP; 1499 A; 605 C; 970 G; 838 T; 0 U; 0 Other;

Query Match 82.0%; Score 16.4; DB 8; Length 3912;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 1172 AAGTAAATGCAGAAACAG 1189

Search completed: January 27, 2006, 22:15:38  
Job time : 194.966 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

QM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:01:27 ; Search time 364.315 Seconds  
(without alignments)  
476.668 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcatagtgggtatcttcc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA Main.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US10F\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	100.0	21	7	US-10-465-205-2
2	21	100.0	21	9	US-10-716-005-2
3	21	100.0	1323	7	US-10-465-205-12
4	21	100.0	1378	7	US-10-465-205-14
5	21	100.0	1379	7	US-10-465-205-7
6	21	100.0	1379	7	US-10-465-205-13
7	21	100.0	1379	7	US-10-465-205-16
8	21	100.0	1384	7	US-10-465-205-8
9	21	100.0	1384	7	US-10-465-205-10
10	21	100.0	1395	7	US-10-465-205-9
11	21	100.0	1390	7	US-10-465-205-11
12	21	100.0	1393	7	US-10-465-205-15
13	21	100.0	1447	7	US-10-465-205-17
14	21	100.0	1734	7	US-10-282-122A-38442
15	21	100.0	1803	7	US-10-465-205-5
16	19.4	92.4	402	8	US-10-472-928-2351
17	19.4	92.4	1287	9	US-10-617-320-1469
18	19.4	92.4	1734	3	US-09-815-242-9183
19	19.4	92.4	1734	7	US-10-282-122A-37804
20	19.4	92.4	8195	2	US-08-961-527-94
21	19.4	92.4	8195	7	US-10-158-844-94
22	19.4	92.4	2162598	8	US-10-472-928-4979
23	18.4	87.6	553	6	US-10-029-386-1376

C 24	18.4	87.6	1084	5	US-10-027-632-10931	Sequence 10931, A
C 25	18.4	87.6	1084	5	US-10-027-632-10932	Sequence 10932, A
C 26	18.4	87.6	1084	5	US-10-027-632-10933	Sequence 10933, A
C 27	18.4	87.6	1084	6	US-10-027-632-10931	Sequence 10931, A
C 28	18.4	87.6	1084	6	US-10-027-632-10932	Sequence 10932, A
C 29	18.4	87.6	1084	6	US-10-027-632-10933	Sequence 10933, A
C 30	17.8	84.8	1830	9	US-10-501-282-5273	Sequence 5273, Ap
C 31	17.8	84.8	1830	9	US-10-501-282-5275	Sequence 5275, Ap
C 32	17.8	84.8	1830	9	US-10-501-282-5277	Sequence 5277, Ap
C 33	17.8	84.8	1830	9	US-10-501-282-5279	Sequence 5279, Ap
C 34	17.8	84.8	2914	6	US-10-094-749-914	Sequence 914, Appl
C 35	17.8	84.8	2914	7	US-10-467-506A-15	Sequence 15, Appl
C 36	17.8	84.8	1754382	9	US-10-501-282-6651	Sequence 6651, Ap
C 37	17.4	82.9	25	8	US-10-719-900-73644	Sequence 73644, A
C 38	16.8	80.0	412	3	US-09-974-300-1665	Sequence 1665, Ap
C 39	16.8	80.0	532	4	US-09-925-065A-30849	Sequence 30849, A
C 40	16.8	80.0	532	4	US-09-925-065A-30850	Sequence 30850, A
C 41	16.8	80.0	616	4	US-09-925-065A-35263	Sequence 35263, A
C 42	16.8	80.0	619	4	US-09-925-065A-511481	Sequence 511481, A
C 43	16.8	80.0	675	4	US-09-925-065A-501042	Sequence 501042, A
C 44	16.8	80.0	1464	7	US-10-335-977-4387	Sequence 4387, Ap
C 45	16.8	80.0	1554	7	US-10-335-977-4388	Sequence 4388, Ap
C 46	16.8	80.0	2547	3	US-09-989-739-18	Sequence 18, Appl
C 47	16.8	80.0	10516	3	US-09-754-468-29	Sequence 29, Appl
C 48	16.8	80.0	51281	3	US-09-997-722-139	Sequence 139, Appl
C 49	16.8	80.0	160921	5	US-10-087-192-1672	Sequence 1672, Ap
C 50	16.4	78.1	520	4	US-09-925-065A-452149	Sequence 452149, A
C 51	16.4	78.1	531	4	US-09-925-065A-392178	Sequence 392178, A
C 52	16.4	78.1	643	4	US-09-925-065A-312506	Sequence 312506, A
C 53	16.4	78.1	647	4	US-09-925-065A-453189	Sequence 453189, A
C 54	16.4	78.1	111331	9	US-10-461-862-101	Sequence 453190, A
C 55	16.4	78.1	164	6	US-10-029-386-24435	Sequence 101, Appl
C 56	16.2	77.1	164	6	US-10-914-037-626	Sequence 24435, A
C 57	16.2	77.1	371	8	US-10-914-037-626	Sequence 626, Appl
C 58	16.2	77.1	434	3	US-09-783-590-9031	Sequence 9031, Ap
C 59	16.2	77.1	520	4	US-09-925-065A-358678	Sequence 358678, A
C 60	16.2	77.1	540	6	US-10-029-386-10735	Sequence 10735, A

## ALIGNMENTS

RESULT 1  
US-10-465-205-2  
; Sequence 2, Application US/10465205  
; Publication NO. US20040014118A1  
; GENERAL INFORMATION:  
; APPLICANT: Uh1, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/465,205  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR APPLICATION NUMBER: US/10/081,923  
; PRIOR FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 21  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide  
US-10-465-205-2

Query Match 100.0%; Score 21; DB 7; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.1;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21

Db 1 TGCATGTATGGGTATCTTCC 21

```
RESULT 2
US-10-716-005-2
; Sequence 2, Application US/10716005
; Publication No. US20050106578A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill III, Franklin R.
; APPLICANT: Aichinger, Christian
; APPLICANT: Reiser, Astrid
; TITLE OF INVENTION: Detection of Group B Streptococcus
; FILE REFERENCE: 07039/460001
; CURRENT APPLICATION NUMBER: US/10/716,005
; CURRENT FILING DATE: 2003-11-18
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-10-716-005-2

Query Match      100.0%; Score 21; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTCC 21
Db 1 TGCATGTATGGTTATCTTCC 21

RESULT 3
US-10-465-205-12/c
; Sequence 12, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptaI sequence from isolate no. 2
US-10-465-205-12

Query Match      100.0%; Score 21; DB 7; Length 1323;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTCC 21
Db 193 TGCATGTATGGTTATCTTCC 173

RESULT 4
US-10-465-205-14/c
; Sequence 14, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptaI sequence from isolate no. 6
US-10-465-205-14

Query Match      100.0%; Score 21; DB 7; Length 1379;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTCC 21
Db 209 TGCATGTATGGTTATCTTCC 189

RESULT 5
US-10-465-205-7/c
; Sequence 7, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptaI sequence from isolate no. 6
US-10-465-205-7

Query Match      100.0%; Score 21; DB 7; Length 1378;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTTATCTTCC 21
Db 209 TGCATGTATGGTTATCTTCC 189

RESULT 6
US-10-465-205-13/c
; Sequence 13, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptaI sequence from isolate no. 6
US-10-465-205-13
```



```
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 4
US-10-465-205-13

Query Match      100.0%; Score 21; DB 7; Length 1379;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 7
US-10-465-205-16/c
; Sequence 16, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 1379
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 1
US-10-465-205-16

Query Match      100.0%; Score 21; DB 7; Length 1379;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 210 TGCATGTATGGGTATCTTCC 190

RESULT 8
US-10-465-205-8/c
; Sequence 8, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 5
US-10-465-205-8

Query Match      100.0%; Score 21; DB 7; Length 1384;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 210 TGCATGTATGGGTATCTTCC 190

RESULT 9
US-10-465-205-10/c
; Sequence 10, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1384
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 8
US-10-465-205-10

Query Match      100.0%; Score 21; DB 7; Length 1384;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 209 TGCATGTATGGGTATCTTCC 189

RESULT 10
US-10-465-205-9/c
; Sequence 9, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptsI sequence from isolate no. 7
US-10-465-205-9

Query Match      100.0%; Score 21; DB 7; Length 1385;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21
    |||||
Db 210 TGCATGTATGGGTATCTTCC 190

RESULT 11
US-10-465-205-11/c
; Sequence 11, Application US/10465205
```

```
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pIsI sequence from isolate no. 9
US-10-465-205-11

Query Match      100.0%; Score 21; DB 7; Length 1390;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TGCATGTATGGGTTATCTTCC 21
Db      224 TGCATGTATGGGTTATCTTCC 204

RESULT 12
US-10-465-205-15/c
; Sequence 15, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 1393
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pIsI sequence from isolate no. 3
US-10-465-205-15

Query Match      100.0%; Score 21; DB 7; Length 1393;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TGCATGTATGGGTTATCTTCC 21
Db      210 TGCATGTATGGGTTATCTTCC 190

RESULT 13
US-10-465-205-17/c
; Sequence 17, Application US/10465205
; Publication No. US20040014118A1
; GENERAL INFORMATION:
; APPLICANT: Uhrl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/465,205
; CURRENT FILING DATE: 2003-06-19
; PRIOR APPLICATION NUMBER: US/10/081,923
; PRIOR FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pIsI sequence from isolate no. 11
US-10-465-205-17

Query Match      100.0%; Score 21; DB 7; Length 1447;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 TGCATGTATGGGTTATCTTCC 21
Db      210 TGCATGTATGGGTTATCTTCC 190

RESULT 14
US-10-282-122A-38442/c
; Sequence 38442, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38442
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Streptococcus pyogenes
; OTHER INFORMATION:
US-10-282-122A-38442

Query Match      100.0%; Score 21; DB 7; Length 1734;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 1 TGCATGATGGTTATCTTCC 21  
Db 377 TGCATGATGGTTATCTTCC 357

RESULT 15

US-10-465-205-5/c  
; Sequence 5, Application US/10465205  
; Publication No. US20040014118A1  
; GENERAL INFORMATION:  
; APPLICANT: Uh1, James R.  
; TITLE OF INVENTION: Detection of Group A Streptococcus  
; FILE REFERENCE: 07039-306001  
; CURRENT APPLICATION NUMBER: US/10/465,205  
; CURRENT FILING DATE: 2003-06-19  
; PRIOR APPLICATION NUMBER: US/10/081,923  
; PRIOR FILING DATE: 2002-07-02  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1803  
; TYPE: DNA  
; ORGANISM: Group A Streptococcus  
; FEATURE:  
; OTHER INFORMATION: pti1 sequence from Oklahoma University M1 strain  
; PUBLICATION INFORMATION:  
; AUTHORS: Ferretti et al.  
; JOURNAL: Proc. Natl. Acad. Sci. USA  
; VOLUME: 98  
; PAGES: 4658-4663  
; DATE: 2001-01-01  
; US-10-465-205-5

Query Match 100.0%; Score 21; DB 7; Length 1803;  
Best Local Similarity 100.0%; Pred. No. 3.5;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGCATGATGGTTATCTTCC 21  
Db 377 TGCATGATGGTTATCTTCC 357

RESULT 16

US-10-472-928-2351/c  
; Sequence 2351, Application US/10472928  
; Publication No. US20050020813A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; TITLE OF INVENTION: THE INSTITUTE FOR GENOMIC RESEARCH  
; FILE REFERENCE: P026926W0  
; CURRENT APPLICATION NUMBER: US/10/472,928  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: GB-0107658.7  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 4979  
; SOFTWARE: SeqWin99, version 1.03  
; SEQ ID NO 2351  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
; US-10-472-928-2351

Query Match 92.4%; Score 19.4; DB 8; Length 402;  
Best Local Similarity 95.2%; Pred. No. 18;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGATGGTTATCTTCC 21  
Db 377 TGCATGATGGTTATCTTCC 357

RESULT 17

US-10-617-320-1469/c  
; Sequence 1469, Application US/10617320  
; Publication No. US20050136404A1  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/617,320  
; FILING DATE: 10-Jul-2003  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 1469:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1287 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...1287  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1469:  
US-10-617-320-1469

Query Match 92.4%; Score 19.4; DB 9; Length 1287;  
Best Local Similarity 95.2%; Pred. No. 21;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGATGGTTATCTTCC 21  
Db 431 TGCATGATGGTTATCTTCC 411

RESULT 18

US-09-815-242-9183/c  
; Sequence 9183, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.

```
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9183
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1734)
US-09-815-242-9183

Query Match          92.4%; Score 19.4; DB 3; Length 1734;
Best Local Similarity 95.2%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
        |||||
Db      377 TGCATGTATGGGTATCTTCC 357

RESULT 19
US-10-282-122A-37804/c
; Sequence 37804, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
```

```
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37804
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-37804

Query Match          92.4%; Score 19.4; DB 7; Length 1734;
Best Local Similarity 95.2%; Pred. No. 22;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
        |||||
Db      377 TGCATGTATGGGTATCTTCC 357

RESULT 20
US-08-961-527-94/c
; Sequence 94, Application US/08961527
; Publication No. US20020032323A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunach
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-94
```

```
Query Match          92.4%; Score 19.4; DB 2; Length 8195;
Best Local Similarity 95.2%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
Db 1384 TGCATGTATGGGTTGTCCTCC 1364

RESULT 21
US-10-158-844-94/c
; Sequence 94, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 94:
US-10-158-844-94

Query Match          92.4%; Score 19.4; DB 7; Length 8195;
Best Local Similarity 95.2%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
Db 1384 TGCATGTATGGGTTGTCCTCC 1364

RESULT 22
US-10-472-928-4979
; Sequence 4979, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
```

```
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4979
; LENGTH: 2162598
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-472-928-4979

Query Match          92.4%; Score 19.4; DB 8; Length 2162598;
Best Local Similarity 95.2%; Pred. No. 49;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 21
Db 1008618 TGCATGTATGGGTTGTCCTCC 1008638

RESULT 23
US-10-029-386-1376
; Sequence 1376, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEOMICA-X-2
; FILE REFERENCE: EXPRESSION ANALYSIS TWO
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1376
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR21 70.0
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: SWISSPROT HIT: O14771, EVALUE 4.00e-02
; OTHER INFORMATION: EST HUMAN HIT: AA228964.1, EVALUE 2.00e-01
; OTHER INFORMATION: NT HIT: AF010322.1, EVALUE 2.10e+00
US-10-029-386-1376

Query Match          87.6%; Score 18.4; DB 6; Length 553;
Best Local Similarity 95.0%; Pred. No. 60;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTTATCTTCC 20
Db 213 TGCATGTATGAGTTATCTTC 232

RESULT 24
US-10-027-632-10931/c
; Sequence 10931, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
```

; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10931  
; LENGTH: 1084  
; TYPE: DNA  
; ORGANISM: Human  
; US-10-027-632-10931

Query Match 87.6%; Score 18.4; DB 5; Length 1084;  
Best Local Similarity 95.0%; Pred. No. 65;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
|||||  
Db 693 TGCATGTATGAGTTATCTTC 674

RESULT 25  
US-10-027-632-10932/c  
; Sequence 10932, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10932  
; LENGTH: 1084  
; TYPE: DNA  
; ORGANISM: Human  
; US-10-027-632-10932

Query Match 87.6%; Score 18.4; DB 5; Length 1084;  
Best Local Similarity 95.0%; Pred. No. 65;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
|||||  
Db 693 TGCATGTATGAGTTATCTTC 674

RESULT 26  
US-10-027-632-10933/c  
; Sequence 10933, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10933  
; LENGTH: 1084  
; TYPE: DNA  
; ORGANISM: Human  
; US-10-027-632-10933

Query Match 87.6%; Score 18.4; DB 5; Length 1084;  
Best Local Similarity 95.0%; Pred. No. 65;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
|||||  
Db 693 TGCATGTATGAGTTATCTTC 674

RESULT 27  
US-10-027-632-10931/c  
; Sequence 10931, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10931  
; LENGTH: 1084  
; TYPE: DNA  
; ORGANISM: Human  
; US-10-027-632-10931

Query Match 87.6%; Score 18.4; DB 6; Length 1084;  
Best Local Similarity 95.0%; Pred. No. 65;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1  TGCATGTATGGTTATCTTC 20
Db      693 TGCATGTATGAGTTATCTTC 674

RESULT 28
US-10-027-632-10932/c
; Sequence 10932, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10932
; LENGTH: 1084
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-10932

Query Match      87.6%; Score 18.4; DB 6; Length 1084;
Best Local Similarity 95.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  TGCATGTATGGTTATCTTC 20
Db      693 TGCATGTATGAGTTATCTTC 674

RESULT 29
US-10-027-632-10933/c
; Sequence 10933, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
```

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10933
; LENGTH: 1084
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-10933

Query Match      87.6%; Score 18.4; DB 6; Length 1084;
Best Local Similarity 95.0%; Pred. No. 65;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      1  TGCATGTATGGTTATCTTC 20
Db      693 TGCATGTATGAGTTATCTTC 674

RESULT 30
US-10-501-282-5273/c
; Sequence 5273, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5273
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (109)..(1827)
US-10-501-282-5273

Query Match      84.8%; Score 17.8; DB 9; Length 1830;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  TGCATGTATGGTTATCTTC 21
Db      485 TGCATGTAGGGTTGTCCTCC 465

RESULT 31
US-10-501-282-5275/c
; Sequence 5275, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
```

```
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5275
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (88)..(1827)
US-10-501-282-5275

Query Match      84.8%; Score 17.8; DB 9; Length 1830;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
         ||||| ||||| ||||| |||||
Db      485 TGCATGTAGGGTGTCTTCC 465

RESULT 32
US-10-501-282-5277/c
; Sequence 5277, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5277
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)..(1827)
US-10-501-282-5277

Query Match      84.8%; Score 17.8; DB 9; Length 1830;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
         ||||| ||||| ||||| |||||
Db      485 TGCATGTAGGGTGTCTTCC 465

RESULT 33
US-10-501-282-5279/c
; Sequence 5279, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCMICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
```

```
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AM100780 L2
; CURRENT APPLICATION NUMBER: US/10/501,282
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5279
; LENGTH: 1830
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1827)
US-10-501-282-5279

Query Match      84.8%; Score 17.8; DB 9; Length 1830;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTCC 21
         ||||| ||||| ||||| |||||
Db      485 TGCATGTAGGGTGTCTTCC 465

RESULT 34
US-10-094-749-914/c
; Sequence 914, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAWATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 914
; LENGTH: 2914
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-914

Query Match      84.8%; Score 17.8; DB 6; Length 2914;
Best Local Similarity 90.5%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```



QY 1 TGCATGATGGGTATCTTCC 21  
 Db 1517 TGCATGATGGGTATCTTCC 1497

RESULT 35  
 US-10-467-506A-15/C  
 ; Sequence 15, Application US/10467506A  
 ; Publication No. US20040170994A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Callen, David Fredrick  
 ; APPLICANT: Kremmidiotis, Gabriel  
 ; APPLICANT: Whitmore, Scott  
 ; APPLICANT: Gardner, Allison  
 ; APPLICANT: Powell, Jason  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR HUMAN TUMOUR SUPPRESSOR GENES  
 ; FILE REFERENCE: 1386-15  
 ; CURRENT APPLICATION NUMBER: US/10/467, 506A  
 ; CURRENT FILING DATE: 2003-12-03  
 ; PRIOR APPLICATION NUMBER: PCT/AU02/00138  
 ; PRIOR FILING DATE: 2002-02-12  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 15  
 ; LENGTH: 2314  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (333)...(1652)  
 US-10-467-506A-15

Query Match 84.8%; Score 17.8; DB 7; Length 2914;  
 Best Local Similarity 90.5%; Pred. No. 1.4e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTATCTTCC 21  
 Db 1517 TGCATGATGGGTATCTTCC 1497

RESULT 36  
 US-10-501-282-6651  
 ; Sequence 6651, Application US/10501282  
 ; Publication No. US20050203280A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MCMICHAEL, JOHN CALHOUN  
 ; APPLICANT: ZAGURSKY, ROBERT JOHN  
 ; APPLICANT: RUSSELL, DAVID PARRISH  
 ; APPLICANT: FLETCHER, LEAH DIANE  
 ; TITLE OF INVENTION: ALLOIOCOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING  
 ; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF  
 ; FILE REFERENCE: AM100780 L2  
 ; CURRENT APPLICATION NUMBER: US/10/501,282  
 ; CURRENT FILING DATE: 2004-07-09  
 ; PRIOR APPLICATION NUMBER: 60/333,777  
 ; PRIOR FILING DATE: 2001-11-29  
 ; PRIOR APPLICATION NUMBER: 60/426,742  
 ; PRIOR FILING DATE: 2002-11-18  
 ; PRIOR APPLICATION NUMBER: PCT/US02/36123  
 ; PRIOR FILING DATE: 2002-11-25  
 ; NUMBER OF SEQ ID NOS: 6653  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 6651  
 ; LENGTH: 1754382  
 ; TYPE: DNA  
 ; ORGANISM: Alloiococcus otitidis  
 US-10-501-282-6651

Query Match 84.8%; Score 17.8; DB 9; Length 1754382;  
 Best Local Similarity 90.5%; Pred. No. 2.9e+02;  
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATGATGGGTATCTTCC 21  
 Db 1394003 TGCATGATGGGTATCTTCC 1394023

RESULT 37  
 US-10-719-900-73644  
 ; Sequence 73644, Application US/10719900  
 ; Publication No. US20050026164A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Xue Mei Zhou  
 ; TITLE OF INVENTION: Methods of Genetic Analysis of Mouse  
 ; FILE REFERENCE: 3528.1  
 ; CURRENT APPLICATION NUMBER: US/10/719,900  
 ; CURRENT FILING DATE: 2003-11-20  
 ; PRIOR APPLICATION NUMBER: 60/427,808  
 ; PRIOR FILING DATE: 2002-11-20  
 ; NUMBER OF SEQ ID NOS: 982914  
 ; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
 ; SEQ ID NO 73644  
 ; LENGTH: 25  
 ; TYPE: DNA  
 ; ORGANISM: Mus musculus  
 US-10-719-900-73644

Query Match 82.9%; Score 17.4; DB 8; Length 25;  
 Best Local Similarity 94.7%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTCC 20  
 Db 5 GCATGTATGGGTATCTTCC 23

RESULT 38  
 US-09-974-300-1665  
 ; Sequence 1665, Application US/09974300  
 ; Patent No. US20020146721A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Berka, Randy M.  
 ; APPLICANT: Clausen, Ib Groth  
 ; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
 ; TITLE OF INVENTION: Expression  
 ; FILE REFERENCE: 10085.500-US  
 ; CURRENT APPLICATION NUMBER: US/09/974,300  
 ; CURRENT FILING DATE: 2001-10-05  
 ; PRIOR APPLICATION NUMBER: 09/680,598  
 ; PRIOR FILING DATE: 2000-10-06  
 ; PRIOR APPLICATION NUMBER: 60/279,526  
 ; PRIOR FILING DATE: 2001-03-27  
 ; NUMBER OF SEQ ID NOS: 8481  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1665  
 ; LENGTH: 412  
 ; TYPE: DNA  
 ; ORGANISM: Bacillus licheniformis  
 US-09-974-300-1665

Query Match 80.0%; Score 16.8; DB 3; Length 412;  
 Best Local Similarity 90.0%; Pred. No. 3.6e+02;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTTCC 21  
 Db 78 GCATGTATGGGTATCTTCC 97

RESULT 39  
 US-09-925-065A-30849/c  
 ; Sequence 30849, Application US/09925065A  
 ; Publication No. US20050228172A9  
 ; GENERAL INFORMATION:

```
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/925,065A
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30849
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-30849

Query Match      80.0%; Score 16.8; DB 4; Length 532;
Best Local Similarity 90.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      468 TGCATGTATGGATTCTTC 449

RESULT 40
US-09-925-065A-30850/c
; Sequence 30850, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/925,065A
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30850
; LENGTH: 532
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-30850

Query Match      80.0%; Score 16.8; DB 4; Length 532;
Best Local Similarity 90.0%; Pred. No. 3.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      468 TGCATGTATGGATTCTTC 449

RESULT 41
US-09-925-065A-355263/c
; Sequence 355263, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/925,065A
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 355263
; LENGTH: 616
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-355263

Query Match      80.0%; Score 16.8; DB 4; Length 616;
Best Local Similarity 90.0%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      206 TGCATGTATGGGTTATCATC 187

RESULT 42
US-09-925-065A-511481
; Sequence 511481, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 09/925,065A
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 511481
; LENGTH: 619
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-511481

Query Match      80.0%; Score 16.8; DB 4; Length 619;
Best Local Similarity 90.0%; Pred. No. 3.8e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      469 TGCATGTATGGGTTATCATC 488
```



```
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...1554
; SEQUENCE DESCRIPTION: SEQ ID NO: 4388;
US-10-335-977-4388

Query Match      80.0%; Score 16.8; DB 7; Length 1554;
Best Local Similarity 90.0%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTC 20
        ||||| ||||| ||||| |||||
Db      1463 TGCATGTATGGGTATCTTC 1444

RESULT 46
US-09-989-739-18/c
; Sequence 18, Application US/09989739
; Publication No. US20030140364A1
; GENERAL INFORMATION:
; APPLICANT: HINCHEY, BRENDAN
; TITLE OF INVENTION: SONG, HEE-SOOK
; TITLE OF INVENTION: MAIZE CYTOPLASMIC GLUTAMINE SYNTHETASE PROMOTER
; FILE REFERENCE: DEKM.177US
; CURRENT APPLICATION NUMBER: US/09/989,739
; CURRENT FILING DATE: 2001-11-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 2547
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
US-09-989-739-18

Query Match      80.0%; Score 16.8; DB 3; Length 2547;
Best Local Similarity 90.0%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTC 20
        ||||| ||||| ||||| |||||
Db      1824 TGCATGTCTGGGTATCTTC 1805

RESULT 47
US-09-754-468-29/c
; Sequence 29, Application US/09754468
; Publication No. US20050192237A1
; GENERAL INFORMATION:
; APPLICANT: Iversen, Patrick L.
; TITLE OF INVENTION: Antisense Antibacterial Cell Division
; TITLE OF INVENTION: Composition and Method
; FILE REFERENCE: 0450-0033.30
; CURRENT APPLICATION NUMBER: US/09/754,468
; CURRENT FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/174,484
; PRIOR FILING DATE: 2000-01-04
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 10516
; TYPE: DNA
; ORGANISM: Helicobacter pylori
US-09-754-468-29

Query Match      80.0%; Score 16.8; DB 3; Length 10516;
Best Local Similarity 90.0%; Pred. No. 5.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTATCTTC 20
```

```
Db      5876 TGCATGTATGGGTATCTTC 5857
        ||||| ||||| ||||| |||||

RESULT 48
US-09-997-722-139/c
; Sequence 139, Application US/09997722
; Publication No. US20040072154A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David
; APPLICANT: Engelhard, Eric
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71171/RMS/DCF
; CURRENT APPLICATION NUMBER: US/09/997,722
; CURRENT FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 301
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 139
; LENGTH: 51281
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3889)..(5506)
; OTHER INFORMATION: "n" at positions 3889 through 5506 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (12248)..(15107)
; OTHER INFORMATION: "n" at positions 12248 through 15107 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34241)..(34260)
; OTHER INFORMATION: "n" at positions 34241 through 34260 can be any base.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (35928)..(35947)
; OTHER INFORMATION: "n" at positions 35928 through 35947 can be any base.
US-09-997-722-139

Query Match      80.0%; Score 16.8; DB 3; Length 51281;
Best Local Similarity 90.0%; Pred. No. 6.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GCATGTATGGGTATCTTC 21
        ||||| ||||| ||||| |||||
Db      37474 GCATGTGGGTATCTTC 37455

RESULT 49
US-10-087-192-1672
; Sequence 1672, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1672
; LENGTH: 160921
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
US-10-087-192-1672

Query Match      80.0%; Score 16.8; DB 5; Length 160921;
Best Local Similarity 90.0%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTC 20
Db      138653 TGCATGTATGTGTTTCTTC 138672

RESULT 50
US-09-925-065A-452149/c
; Sequence 452149, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 452149
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-452149

Query Match      78.1%; Score 16.4; DB 4; Length 520;
Best Local Similarity 94.4%; Pred. No. 5.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      434 TCCATGTATGGGTTATCT 417

RESULT 51
US-09-925-065A-392178/c
; Sequence 392178, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 392178
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-392178

Query Match      78.1%; Score 16.4; DB 4; Length 531;
Best Local Similarity 94.4%; Pred. No. 5.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      435 TCCATGTATGGGTTATCT 418

RESULT 52
US-09-925-065A-312506/c
; Sequence 312506, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 312506
; LENGTH: 643
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-312506

Query Match      78.1%; Score 16.4; DB 4; Length 643;
Best Local Similarity 94.4%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      446 TGCATGTCTGGGTTATCT 429

RESULT 53
US-09-925-065A-453189/c
; Sequence 453189, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 453189/c
; Length: 643
; Type: DNA
; Organism: Homo sapiens
US-09-925-065A-453189/c

Query Match      78.1%; Score 16.4; DB 4; Length 643;
Best Local Similarity 94.4%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      446 TGCATGTCTGGGTTATCT 429

RESULT 54
US-09-925-065A-453189/c
; Sequence 453189, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 453189/c
; Length: 643
; Type: DNA
; Organism: Homo sapiens
US-09-925-065A-453189/c

Query Match      78.1%; Score 16.4; DB 4; Length 643;
Best Local Similarity 94.4%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      446 TGCATGTCTGGGTTATCT 429

RESULT 55
US-09-925-065A-453189/c
; Sequence 453189, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 453189/c
; Length: 643
; Type: DNA
; Organism: Homo sapiens
US-09-925-065A-453189/c

Query Match      78.1%; Score 16.4; DB 4; Length 643;
Best Local Similarity 94.4%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      446 TGCATGTCTGGGTTATCT 429
```

; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 453189  
; LENGTH: 647  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-453189

Query Match 78.1%; Score 16.4; DB 4; Length 647;  
Best Local Similarity 94.4%; Pred. No. 6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCT 18  
||||| |||||||  
DB 446 TGCATGTCTGGGTTATCT 429

RESULT 54  
US-09-925-065A-453190/c  
; Sequence 453190, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 453190  
; LENGTH: 647  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-453190

Query Match 78.1%; Score 16.4; DB 4; Length 647;  
Best Local Similarity 94.4%; Pred. No. 6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCT 18  
||||| |||||||  
DB 446 TGCATGTCTGGGTTATCT 429

RESULT 55  
US-10-461-862-101/c  
; Sequence 101, Application US/10461862  
; Publication No. US20050090434A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer  
; FILE REFERENCE: 529452001800  
; CURRENT APPLICATION NUMBER: US/10/461,862  
; CURRENT FILING DATE: 2003-06-13  
; NUMBER OF SEQ ID NOS: 184  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 101  
; LENGTH: 111331  
; TYPE: DNA  
; ORGANISM: Mus musculus

; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(111331)  
; OTHER INFORMATION: n = A,T,C or G  
US-10-461-862-101

Query Match 78.1%; Score 16.4; DB 9; Length 111331;  
Best Local Similarity 94.4%; Pred. No. 1.1e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTTATCTT 19  
||||| |||||||  
DB 21925 GCATGTATGGGTTATCTT 21908

RESULT 56  
US-10-029-386-24435  
; Sequence 24435, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 24435  
; LENGTH: 164  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO CHR16.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
; OTHER INFORMATION: NT HIT: AF144387.1, EVALUATE 3.60e-02  
; OTHER INFORMATION: SWISSPROT HIT: O28578, EVALUATE 4.40e+00  
US-10-029-386-24435

Query Match 77.1%; Score 16.2; DB 6; Length 164;  
Best Local Similarity 85.7%; Pred. No. 6.5e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
||||| |||||||  
DB 104 TTCTGTATGGATTATCTTCC 124

RESULT 57  
US-10-914-037-626/c  
; Sequence 626, Application US/10914037  
; Publication No. US20050003444A1  
; GENERAL INFORMATION:  
; APPLICANT: Nehls, Michael  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: NOVEL HUMAN POLYNUCLEOTIDES AND THE  
; FILE REFERENCE: 8535-0029-999  
; CURRENT APPLICATION NUMBER: US/10/914,037  
; CURRENT FILING DATE: 2004-08-06  
; PRIOR APPLICATION NUMBER: US/09/428,674  
; PRIOR FILING DATE: 1999-10-27  
; PRIOR APPLICATION NUMBER: US 60/106,442  
; PRIOR FILING DATE: 1998-10-30  
; NUMBER OF SEQ ID NOS: 1008  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 626  
; LENGTH: 371

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-914-037-626

Query Match      77.1%; Score 16.2; DB 8; Length 371;
Best Local Similarity 85.7%; Pred. No. 7.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGCATGTATGGTATCTTCC 21
Db 70 TGCATGTATCTGTGTTCTCC 50

RESULT 58
US-09-783-590-9031/c
; Sequence 9031, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9031
; LENGTH: 494
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (42)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (200)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (205)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (214)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (216)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (254)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (266)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (268)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (269)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (272)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (284)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (286)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (295)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (305)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (306)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (312)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (316)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (320)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (326)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (332)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (340)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (350)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (370)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (372)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (377)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (386)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (395)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (399)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (428)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (429)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (436)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (445)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (446)
; OTHER INFORMATION: n equals a,t,g, or c
```

; NAME/KEY: misc feature  
; LOCATION: (458)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (469)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (475)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (477)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (478)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (483)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc feature  
; LOCATION: (484)  
; OTHER INFORMATION: n equals a,t,g, or c  
; US-09-783-590-9031

Query Match 77.1%; Score 16.2; DB 3; Length 494;  
Best Local Similarity 85.7%; Pred. No. 7.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
||| ||||| |||||  
Db 159 TGGATATATGGTTTCTTCC 139

## RESULT 59

US-09-925-065A-358678/c  
; Sequence 358678, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925.065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 358678  
; LENGTH: 520  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-358678

Query Match 77.1%; Score 16.2; DB 4; Length 520;  
Best Local Similarity 85.7%; Pred. No. 7.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
||| ||||| |||||  
Db 25 TGCATGTATGGCTTCTTCC 5

## RESULT 60

US-10-029-386-10735  
; Sequence 10735, Application US/10029386

; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 10735  
; LENGTH: 540  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO CHR16.3  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
; OTHER INFORMATION: SWISSPROT HIT: P09065, EVALUATE 3.90e+00  
; OTHER INFORMATION: NT HIT: A8004835.1, EVALUATE 1.30e-01  
; OTHER INFORMATION: EST\_HUMAN HIT: AA131046.1, EVALUATE 3.00e+00  
US-10-029-386-10735

Query Match 77.1%; Score 16.2; DB 6; Length 540;  
Best Local Similarity 85.7%; Pred. No. 7.4e+02;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGTTATCTTCC 21  
||| ||||| |||||  
Db 461 TTCTGTATGGATTATCTTCC 481

Search completed: January 28, 2006, 01:45:35  
Job time : 371.315 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:15:53 ; Search time 211.652 Seconds  
(without alignments)  
82.419 Million cell updates/sec

Title: US-10-716-005-2

Perfect score: 21

Sequence: 1 tgcattatgggttatcttcc 21

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6059551 seqs, 415333918 residues

Total number of hits satisfying chosen parameters: 12119102

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA New:  
1: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.2	77.1	25	8	US-11-121-849-323547 Sequence 323547,
2	16.2	77.1	811	7	US-10-750-185-33099 Sequence 33099, A
3	16.2	77.1	811	7	US-10-750-623-33099 Sequence 33099, A
4	16.2	77.1	1712	7	US-10-750-185-57490 Sequence 57490, A
5	16.2	77.1	1712	7	US-10-750-623-57490 Sequence 57490, A
6	16.2	77.1	4975	7	US-10-750-185-37365 Sequence 37365, A
7	16.2	77.1	4975	7	US-10-750-623-37365 Sequence 37365, A
8	15.8	75.2	201	7	US-10-995-561-20395 Sequence 20395, A
9	15.8	75.2	600	7	US-10-750-185-2992 Sequence 2992, Ap
10	15.8	75.2	600	7	US-10-750-623-2992 Sequence 2992, Ap
11	15.8	75.2	1721	8	US-11-010-239-33 Sequence 33, Appl
12	15.8	75.2	380749	7	US-10-995-561-13216 Sequence 13216, A
13	15.8	75.2	1125000	7	US-10-995-561-13286 Sequence 13286, A
14	15.2	72.4	94	7	US-10-310-914A-19647 Sequence 19647, A
15	15.2	72.4	1094	7	US-10-750-185-55160 Sequence 55160, A
16	15.2	72.4	1094	7	US-10-750-623-55160 Sequence 55160, A
17	15.2	72.4	1406	7	US-10-750-185-55478 Sequence 55478, A
18	15.2	72.4	1406	7	US-10-750-623-55478 Sequence 55478, A
19	15.2	72.4	1789	7	US-10-750-185-45422 Sequence 45422, A
20	15.2	72.4	1789	7	US-10-750-623-45422 Sequence 45422, A
21	15.2	72.4	2234	7	US-10-750-185-56274 Sequence 56274, A
22	15.2	72.4	2234	7	US-10-750-623-56274 Sequence 56274, A

C 23	15.2	72.4	2382	7	US-10-750-185-59755	Sequence 59755, A
C 24	15.2	72.4	2382	7	US-10-750-623-59755	Sequence 59755, A
C 25	15.2	72.4	96583	8	US-11-117-187-203	Sequence 203, App
C 26	15.2	72.4	203467	8	US-11-121-086-50	Sequence 50, Appl
C 27	14.8	70.5	19	9	US-11-101-244-89198	Sequence 89198, A
C 28	14.8	70.5	19	10	US-11-083-788-89198	Sequence 89198, A
C 29	14.8	70.5	25	8	US-11-121-849-209933	Sequence 209933, A
C 30	14.8	70.5	25	8	US-11-121-849-273497	Sequence 273497, A
C 31	14.8	70.5	201	7	US-10-995-561-25366	Sequence 25366, A
C 32	14.8	70.5	774	7	US-10-750-185-42892	Sequence 42892, A
C 33	14.8	70.5	774	7	US-10-750-623-42892	Sequence 42892, A
C 34	14.8	70.5	982	7	US-10-750-185-62244	Sequence 62244, A
C 35	14.8	70.5	982	7	US-10-750-623-62244	Sequence 62244, A
C 36	14.8	70.5	999	6	US-10-981-334-45	Sequence 45, Appl
C 37	14.8	70.5	1338	7	US-10-750-185-43249	Sequence 43249, A
C 38	14.8	70.5	1338	7	US-10-750-623-43249	Sequence 43249, A
C 39	14.8	70.5	1560	7	US-10-750-185-64265	Sequence 64265, A
C 40	14.8	70.5	1560	7	US-10-750-623-64265	Sequence 64265, A
C 41	14.8	70.5	1784	7	US-10-750-185-29709	Sequence 29709, A
C 42	14.8	70.5	1784	7	US-10-750-623-29709	Sequence 29709, A
C 43	14.8	70.5	4689	8	US-11-136-527-2197	Sequence 2197, Ap
C 44	14.8	70.5	47444	7	US-10-995-561-13354	Sequence 13354, A
C 45	14.8	70.5	159146	8	US-11-121-086-49	Sequence 49, Appl
C 46	14.8	70.5	199130	7	US-10-995-561-13233	Sequence 13233, A
C 47	14.6	69.5	25	8	US-11-121-849-6595	Sequence 6595, Ap
C 48	14.6	69.5	201	8	US-11-124-368A-8804	Sequence 8804, Ap
C 49	14.6	69.5	868	7	US-10-793-626-1891	Sequence 1891, Ap
C 50	14.6	69.5	912	7	US-10-793-626-1139	Sequence 1139, Ap
C 51	14.6	69.5	1080	7	US-10-793-626-2741	Sequence 2741, Ap
C 52	14.6	69.5	1145	7	US-10-750-185-62709	Sequence 62709, A
C 53	14.6	69.5	1145	7	US-10-750-623-62709	Sequence 62709, A
C 54	14.6	69.5	1208	7	US-10-750-185-62508	Sequence 62508, A
C 55	14.6	69.5	1208	7	US-10-750-623-62508	Sequence 62508, A
C 56	14.6	69.5	1275	7	US-10-750-185-29432	Sequence 29432, A
C 57	14.6	69.5	1275	7	US-10-750-623-29432	Sequence 29432, A
C 58	14.6	69.5	1353	7	US-10-750-185-62216	Sequence 62216, A
C 59	14.6	69.5	1353	7	US-10-750-623-62216	Sequence 62216, A
C 60	14.6	69.5	1420	7	US-10-512-109-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1  
US-11-121-849-323547  
; Sequence 323547, Application US/11121849  
; Publication No. US20050272080A1  
; GENERAL INFORMATION:  
; APPLICANT: John Palma  
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded  
; TITLE OF INVENTION: Microarrays  
; FILE REFERENCE: 3684.1  
; CURRENT APPLICATION NUMBER: US/11/121.849  
; CURRENT FILING DATE: 2005-05-03  
; PRIOR APPLICATION NUMBER: 60/567,949  
; PRIOR FILING DATE: 2004-05-03  
; NUMBER OF SEQ ID NOS: 673904  
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1  
; SEQ ID NO 323547  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Homo sapien  
US-11-121-849-323547

Query Match 77.1%; Score 16.2; DB 8; Length 25;  
Best Local Similarity 85.7%; Pred. No. 26;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 TGCATGTATGGGTATCTTCC 21  
Db 5 TGAATGTATGGGTCTTTTCC 25

RESULT 2  
US-10-750-185-33099  
; Sequence 33099, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 33099  
; LENGTH: 811  
; TYPE: DNA  
; ORGANISM: Bovine 19866880817752  
US-10-750-185-33099

Query Match 77.1%; Score 16.2; DB 7; Length 811;  
Best Local Similarity 85.7%; Pred. No. 49;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 632 TGCATGGATGGGTTTCTACC 652

RESULT 3  
US-10-750-623-33099  
; Sequence 33099, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 33099  
; LENGTH: 811  
; TYPE: DNA  
; ORGANISM: Bovine 19866880817752  
US-10-750-623-33099

Query Match 77.1%; Score 16.2; DB 7; Length 811;  
Best Local Similarity 85.7%; Pred. No. 49;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 632 TGCATGGATGGGTTTCTACC 652

RESULT 4  
US-10-750-185-57490  
; Sequence 57490, Application US/10750185

; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 57490  
; LENGTH: 1712  
; TYPE: DNA  
; ORGANISM: Bovine 19866880874779  
US-10-750-185-57490

Query Match 77.1%; Score 16.2; DB 7; Length 1712;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 1314 TGCATATGTGGGTGATCTTCC 1334

RESULT 5  
US-10-750-623-57490  
; Sequence 57490, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 57490  
; LENGTH: 1712  
; TYPE: DNA  
; ORGANISM: Bovine 19866880874779  
US-10-750-623-57490

Query Match 77.1%; Score 16.2; DB 7; Length 1712;  
Best Local Similarity 85.7%; Pred. No. 56;  
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 1314 TGCATATGTGGGTGATCTTCC 1334

RESULT 6  
US-10-750-185-37365/c  
; Sequence 37365, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.

```
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37365
; LENGTH: 4975
; TYPE: DNA
; ORGANISM: Bovine 19866879992949
US-10-750-185-37365

Query Match      77.1%; Score 16.2; DB 7; Length 4975;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 TGCATGTATGGGTATCTTC 21
Db      1760 TGCATGTATGGGTATCTTC 1740

RESULT 7
US-10-750-623-37365/c
; Sequence 37365, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37365
; LENGTH: 4975
; TYPE: DNA
; ORGANISM: Bovine 19866879992949
US-10-750-623-37365

Query Match      77.1%; Score 16.2; DB 7; Length 4975;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy      1 TGCATGTATGGGTATCTTC 21
Db      1760 TGCATGTATGGGTATCTTC 1740

RESULT 8
US-10-995-561-20395/c
; Sequence 20395, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF

; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20395
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-20395

Query Match      75.2%; Score 15.8; DB 7; Length 201;
Best Local Similarity 89.5%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 TGCATGTATGGGTATCTT 19
Db      57 TGCATATATGAGTTATCTT 39

RESULT 9
US-10-750-185-2992/c
; Sequence 2992, Application US/10750185
; Publication No. US2005028603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2992
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Bovine MMBT13234
US-10-750-185-2992

Query Match      75.2%; Score 15.8; DB 7; Length 600;
Best Local Similarity 89.5%; Pred. No. 75;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 TGCATGTATGGGTATCTT 19
Db      425 TGTATATATGGGTATCTT 407

RESULT 10
US-10-750-623-2992/c
; Sequence 2992, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
```



Best Local Similarity 40.0%; Pred. No. 1.1e+02; Mismatches 9; Indels 3; Gaps 0;  
Matches 8; Conservative

QY 1 TGCATGTATGGGTATCTTC 20  
Db 64 UGGAUUUUUGGUUAUCUUC 83

RESULT 15  
US-10-750-185-55160/c  
; Sequence 55160, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 55160  
; LENGTH: 1094  
; TYPE: DNA  
; ORGANISM: Bovine 19866881169381  
US-10-750-185-55160

Query Match : 72.4%; Score 15.2; DB 7; Length 1094;  
Best Local Similarity 85.0%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21  
Db 611 GCAAGTATAGGTTTCTTC 592

RESULT 16  
US-10-750-623-55160/c  
; Sequence 55160, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 55160  
; LENGTH: 1094  
; TYPE: DNA  
; ORGANISM: Bovine 19866881169381  
US-10-750-623-55160

Query Match 72.4%; Score 15.2; DB 7; Length 1094;  
Best Local Similarity 85.0%; Pred. No. 1.7e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 2 GCATGTATGGGTATCTTC 21  
Db 611 GCAAGTATAGGTTTCTTC 592

RESULT 17  
US-10-750-185-55478  
; Sequence 55478, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 55478  
; LENGTH: 1406  
; TYPE: DNA  
; ORGANISM: Bovine 19866880671997  
US-10-750-185-55478

Query Match 72.4%; Score 15.2; DB 7; Length 1406;  
Best Local Similarity 85.0%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
Db 1082 TGAATGTATGTGTATCATC 1101

RESULT 18  
US-10-750-623-55478  
; Sequence 55478, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 55478  
; LENGTH: 1406  
; TYPE: DNA  
; ORGANISM: Bovine 19866880671997  
US-10-750-623-55478

Query Match 72.4%; Score 15.2; DB 7; Length 1406;  
Best Local Similarity 85.0%; Pred. No. 1.8e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 1 TGCATGTATGGGTATCTTC 20  
Db 1082 TGAATGTATGTGTATCATC 1101

```
US-10-750-185-45422
; Sequence 45422, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45422
; LENGTH: 1789
; TYPE: DNA
; ORGANISM: Bovine 19866880353593
US-10-750-185-45422

Query Match          72.4%; Score 15.2; DB 7; Length 1789;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GCATGATGGGTATCTTCC 21
| | | | | | | | | | | | | | |
Db      619 GAATGATGGGTATCTTCC 638

RESULT 20
US-10-750-623-45422
; Sequence 45422, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 45422
; LENGTH: 1789
; TYPE: DNA
; ORGANISM: Bovine 19866880353593
US-10-750-623-45422

Query Match          72.4%; Score 15.2; DB 7; Length 1789;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2 GCATGATGGGTATCTTCC 21
| | | | | | | | | | | | | | |
Db      619 GAATGATGGGTATCTTCC 638

US-10-750-185-56274
; Sequence 56274, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56274
; LENGTH: 2234
; TYPE: DNA
; ORGANISM: Bovine 19866880566031
US-10-750-185-56274

Query Match          72.4%; Score 15.2; DB 7; Length 2234;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGCATGATGGGTATCTTC 20
| | | | | | | | | | | | | | |
Db      816 TGCATATATGGGTCTTATTC 835

RESULT 22
US-10-750-623-56274
; Sequence 56274, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; CURRENT FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56274
; LENGTH: 2234
; TYPE: DNA
; ORGANISM: Bovine 19866880566031
US-10-750-623-56274

Query Match          72.4%; Score 15.2; DB 7; Length 2234;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1 TGCATGATGGGTATCTTC 20
| | | | | | | | | | | | | | |
Db      816 TGCATATATGGGTCTTATTC 835

RESULT 23
US-10-750-185-59755/c
; Sequence 59755, Application US/10750185
; Publication No. US20050260603A1
```

```

; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59755
; LENGTH: 2382
; TYPE: DNA
; ORGANISM: Bovine 19866880423655
US-10-750-185-59755

```

```

Query Match 72.4%; Score 15.2; DB 7; Length 2382;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 GCATGTATGGTTATCTTCC 21
Db 1168 GCATTAATGGTTGTTCTTCC 1149

```

```

RESULT 24
US-10-750-623-59755/c
; Sequence 59755, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59755
; LENGTH: 2382
; TYPE: DNA
; ORGANISM: Bovine 19866880423655
US-10-750-623-59755

```

```

Query Match 72.4%; Score 15.2; DB 7; Length 2382;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 GCATGTATGGTTATCTTCC 21
Db 1168 GCATTAATGGTTGTTCTTCC 1149

```

```

RESULT 25
US-11-117-187-203
; Sequence 203, Application US/11117187
; Publication No. US20050266560A1
; GENERAL INFORMATION:
; APPLICANT: PREUSS, DAPHNE
; APPLICANT: COPENHAVER, GREGORY

```

```

; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS
; FILE REFERENCE: ARCD:309US
; CURRENT APPLICATION NUMBER: US/11/117,187
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: US/09/531,120
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/125,219
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 203
; LENGTH: 96583
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-11-117-187-203

```

```

Query Match 72.4%; Score 15.2; DB 8; Length 96583;
Best Local Similarity 85.0%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 TGCATGTATGGTTATCTTCC 20
Db 38324 TGCATGTATGGTTTCTTCC 38343

```

```

RESULT 26
US-11-121-086-50/c
; Sequence 50, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 50
; LENGTH: 203467
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-50

```

```

Query Match 72.4%; Score 15.2; DB 8; Length 203467;
Best Local Similarity 85.0%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 2 GCATGTATGGTTATCTTCC 21
Db 90768 GCATGTCTTAGTTATCTTCC 90749

```

```

RESULT 27
US-11-101-244-89198/c
; Sequence 89198, Application US/11101244
; Publication No. US20050246794A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/101,244
; CURRENT FILING DATE: 2005-04-07
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137

```

```
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 89198
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-101-244-89198

Query Match      70.5%; Score 14.8; DB 9; Length 19;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 ATGTATGGGTTATCTTCC 21
      ||||| ||||| |||||
Db      18 ATGCTGTGTTATCTTCC 1

RESULT 28
US-11-083-784-89198/c
; Sequence 89198, Application US/11083784
; Publication No. US20050245475A1
; GENERAL INFORMATION:
; APPLICANT: Dharmacon, Inc.
; APPLICANT: Khvorova, Anastasia
; APPLICANT: Reynolds, Angela
; APPLICANT: Leake, Devin
; APPLICANT: Marshall, William
; APPLICANT: Scaringe, Stephen
; TITLE OF INVENTION: Functional and Hyperfunctional siRNA
; FILE REFERENCE: 13499US
; CURRENT APPLICATION NUMBER: US/11/083,784
; CURRENT FILING DATE: 2005-03-18
; PRIOR APPLICATION NUMBER: US/10/714,333
; PRIOR FILING DATE: 2003-11-14
; PRIOR APPLICATION NUMBER: 60/502,050
; PRIOR FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: 60/426,137
; PRIOR FILING DATE: 2002-11-14
; NUMBER OF SEQ ID NOS: 1591911
; SOFTWARE: Proprietary
; SEQ ID NO 89198
; LENGTH: 19
; TYPE: RNA
; ORGANISM: Homo sapiens
US-11-083-784-89198

Query Match      70.5%; Score 14.8; DB 10; Length 19;
Best Local Similarity 88.9%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 ATGTATGGGTTATCTTCC 21
      ||||| ||||| |||||
Db      18 ATGCTGTGTTATCTTCC 1

RESULT 29
US-11-121-849-200933/c
; Sequence 200933, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; TITLE OF INVENTION: Microarrays
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 200933
; LENGTH: 25
```

```
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-200933

Query Match      70.5%; Score 14.8; DB 8; Length 25;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
      ||||| ||||| |||||
Db      24 TGCATGTATGGGTTGCT 7

RESULT 30
US-11-121-849-273497
; Sequence 273497, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded S
; TITLE OF INVENTION: Microarrays
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 273497
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapien
US-11-121-849-273497

Query Match      70.5%; Score 14.8; DB 8; Length 25;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 ATGTATGGGTTATCTTCC 21
      ||||| ||||| |||||
Db      6 ATGCTGTGATTATCTTCC 23

RESULT 31
US-10-995-561-25366/c
; Sequence 25366, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25366
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-25366

Query Match      70.5%; Score 14.8; DB 7; Length 201;
Best Local Similarity 88.9%; Pred. No. 2e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 ATGTATGGGTTATCTTCC 21
      ||||| ||||| |||||
Db      121 ATGCATGGGTTAGCTTCC 104

RESULT 32
```



US-10-750-185-42892/c  
; Sequence 42892, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 42892  
; LENGTH: 774  
; TYPE: DNA  
; ORGANISM: Bovine 19866880852756  
US-10-750-185-42892

Query Match 70.5%; Score 14.8; DB 7; Length 774;  
Best Local Similarity 88.9%; Pred. No. 2.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTT 19  
||||| ||||| |||||  
Db 624 GCATGTGGGTTTCTT 607

RESULT 33  
US-10-750-623-42892/c  
; Sequence 42892, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 42892  
; LENGTH: 774  
; TYPE: DNA  
; ORGANISM: Bovine 19866880852756  
US-10-750-623-42892

Query Match 70.5%; Score 14.8; DB 7; Length 774;  
Best Local Similarity 88.9%; Pred. No. 2.6e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCATGTATGGGTATCTT 19  
||||| ||||| |||||  
Db 624 GCATGTGGGTTTCTT 607

RESULT 34  
US-10-750-185-62244  
; Sequence 62244, Application US/10750185  
; Publication No. US20050260603A1

; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 62244  
; LENGTH: 982  
; TYPE: DNA  
; ORGANISM: Bovine 1986688112236  
US-10-750-185-62244

Query Match 70.5%; Score 14.8; DB 7; Length 982;  
Best Local Similarity 88.9%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTCC 21  
||||| ||||| |||||  
Db 156 ATGTATGTATATCTTCC 173

RESULT 35  
US-10-750-623-62244  
; Sequence 62244, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 62244  
; LENGTH: 982  
; TYPE: DNA  
; ORGANISM: Bovine 1986688112236  
US-10-750-623-62244

Query Match 70.5%; Score 14.8; DB 7; Length 982;  
Best Local Similarity 88.9%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTCC 21  
||||| ||||| |||||  
Db 156 ATGTATGTATATCTTCC 173

RESULT 36  
US-10-981-334-45  
; Sequence 45, Application US/10981334  
; Publication No. US2006008816A1  
; GENERAL INFORMATION:  
; APPLICANT: LU, Yu-Ping et al.  
; TITLE OF INVENTION: PROMOTER, PROMOTER CONTROL ELEMENTS, AND COMBINATIONS, AND USES

;/ TITLE OF INVENTION: THEREOF  
;/ FILE REFERENCE: 2750-1561PUS2  
;/ CURRENT APPLICATION NUMBER: US/10/981,334  
;/ CURRENT FILING DATE: 2004-11-04  
;/ NUMBER OF SEQ ID NOS: 68  
;/ SOFTWARE: PatentIn version 3.3  
;/ SEQ ID NO 45  
;/ LENGTH: 999  
;/ TYPE: DNA  
;/ ORGANISM: Unknown  
;/ FEATURE:  
;/ OTHER INFORMATION: Promoter and/or promoter control element identified from  
;/ OTHER INFORMATION: Arabidopsis thaliana or Oryza sativa.  
US-10-981-334-45

Query Match 70.5%; Score 14.8; DB 6; Length 999;  
Best Local Similarity 88.9%; Pred. No. 2.7e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTATCTTC 20  
||| ||||| |||||  
Db 848 CATTATGGATTATCTTC 865

RESULT 37  
US-10-750-185-43249  
;/ Sequence 43249, Application US/10750185  
;/ Publication No. US20050260603A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: MMI GENOMICS, INC.  
;/ APPLICANT: DENISE, Sue K.  
;/ APPLICANT: KERR, Richard  
;/ APPLICANT: ROSENFELD, David  
;/ APPLICANT: HOLM, Tom  
;/ APPLICANT: BATES, Stephen  
;/ APPLICANT: FANTIN, Dennis  
;/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
;/ FILE REFERENCE: MM1100-2  
;/ CURRENT APPLICATION NUMBER: US/10/750,185  
;/ CURRENT FILING DATE: 2003-12-31  
;/ PRIOR APPLICATION NUMBER: US 60/437,482  
;/ PRIOR FILING DATE: 2002-12-31  
;/ NUMBER OF SEQ ID NOS: 64922  
;/ SOFTWARE: PatentIn version 3.1  
;/ SEQ ID NO 43249  
;/ LENGTH: 1338  
;/ TYPE: DNA  
;/ ORGANISM: Bovine 19866881139126  
US-10-750-185-43249

Query Match 70.5%; Score 14.8; DB 7; Length 1338;  
Best Local Similarity 88.9%; Pred. No. 2.8e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTC 21  
||| ||||| |||||  
Db 205 ATGCATGGGTATTTCC 222

RESULT 38  
US-10-750-623-43249  
;/ Sequence 43249, Application US/10750623  
;/ Publication No. US20050287531A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: MMI GENOMICS, INC.  
;/ APPLICANT: DENISE, Sue K.  
;/ APPLICANT: KERR, Richard  
;/ APPLICANT: ROSENFELD, David  
;/ APPLICANT: HOLM, Tom  
;/ APPLICANT: BATES, Stephen  
;/ APPLICANT: FANTIN, Dennis  
;/ TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
;/ FILE REFERENCE: MM1100-1

;/ CURRENT APPLICATION NUMBER: US/10/750,623  
;/ CURRENT FILING DATE: 2003-12-31  
;/ PRIOR APPLICATION NUMBER: US 60/437,482  
;/ PRIOR FILING DATE: 2002-12-31  
;/ NUMBER OF SEQ ID NOS: 64922  
;/ SOFTWARE: PatentIn version 3.1  
;/ SEQ ID NO 43249  
;/ LENGTH: 1338  
;/ TYPE: DNA  
;/ ORGANISM: Bovine 19866881139126  
US-10-750-623-43249

Query Match 70.5%; Score 14.8; DB 7; Length 1338;  
Best Local Similarity 88.9%; Pred. No. 2.8e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATGTATGGGTATCTTC 21  
||| ||||| |||||  
Db 205 ATGCATGGGTATTTCC 222

RESULT 39  
US-10-750-185-64265  
;/ Sequence 64265, Application US/10750185  
;/ Publication No. US20050260603A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: MMI GENOMICS, INC.  
;/ APPLICANT: DENISE, Sue K.  
;/ APPLICANT: KERR, Richard  
;/ APPLICANT: ROSENFELD, David  
;/ APPLICANT: HOLM, Tom  
;/ APPLICANT: BATES, Stephen  
;/ APPLICANT: FANTIN, Dennis  
;/ TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
;/ FILE REFERENCE: MM1100-2  
;/ CURRENT APPLICATION NUMBER: US/10/750,185  
;/ CURRENT FILING DATE: 2003-12-31  
;/ PRIOR APPLICATION NUMBER: US 60/437,482  
;/ PRIOR FILING DATE: 2002-12-31  
;/ NUMBER OF SEQ ID NOS: 64922  
;/ SOFTWARE: PatentIn version 3.1  
;/ SEQ ID NO 64265  
;/ LENGTH: 1560  
;/ TYPE: DNA  
;/ ORGANISM: Bovine 19866881415571  
US-10-750-185-64265

Query Match 70.5%; Score 14.8; DB 7; Length 1560;  
Best Local Similarity 88.9%; Pred. No. 2.9e+02;  
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 CATGTATGGGTATCTTC 20  
||| ||||| |||||  
Db 1129 CATTATGGGTATCTAC 1146

RESULT 40  
US-10-750-623-64265  
;/ Sequence 64265, Application US/10750623  
;/ Publication No. US20050287531A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: MMI GENOMICS, INC.  
;/ APPLICANT: DENISE, Sue K.  
;/ APPLICANT: KERR, Richard  
;/ APPLICANT: ROSENFELD, David  
;/ APPLICANT: HOLM, Tom  
;/ APPLICANT: BATES, Stephen  
;/ APPLICANT: FANTIN, Dennis  
;/ TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
;/ FILE REFERENCE: MM1100-1  
;/ CURRENT APPLICATION NUMBER: US/10/750,623  
;/ CURRENT FILING DATE: 2003-12-31  
;/ PRIOR APPLICATION NUMBER: US 60/437,482



```
QY      4 ATGATATGGGTTATCTTCC 21
Db      2089 ATTATATGGTTATCTTCC 2072

RESULT 45
US-11-121-086-49/c
; Sequence 49, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 49
; LENGTH: 159146
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-49

Query Match      70.5%; Score 14.8; DB 8; Length 159146;
Best Local Similarity 88.9%; Pred. No. 6.4e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCT 18
Db      98113 TGAATGTATGGGTTGTCT 98096

RESULT 46
US-10-995-561-13233/c
; Sequence 13233, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13233
; LENGTH: 199130
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(199130)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13233

Query Match      70.5%; Score 14.8; DB 7; Length 199130;
Best Local Similarity 88.9%; Pred. No. 6.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      4 ATGATATGGGTTATCTTCC 21
Db      190002 ATGCATGGTTAGCTTCC 189985

RESULT 47
US-11-121-849-6595/c
; Sequence 6595, Application US/11121849
; Publication No. US20050272080A1
; GENERAL INFORMATION:
; APPLICANT: JOHN PALMA
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded ;
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 6595
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-849-6595

Query Match      69.5%; Score 14.6; DB 8; Length 25;
Best Local Similarity 81.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
Db      21 TACATGCATGGGTTTCTGCC 1

RESULT 48
US-11-124-368A-8804/c
; Sequence 8804, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8804
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-8804

Query Match      69.5%; Score 14.6; DB 8; Length 201;
Best Local Similarity 81.0%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
Db      187 TGAATGCATGGGTTTGTCTCC 167

RESULT 49
US-10-793-626-1891/c
; Sequence 1891, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
```

```
; APPLICANT: John Palma
; TITLE OF INVENTION: Methods of Genetic Analysis of Formalin Fixed Paraffin Embedded ;
; FILE REFERENCE: 3684.1
; CURRENT APPLICATION NUMBER: US/11/121,849
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: 60/567,949
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 673904
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 6595
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-849-6595
```

```
Query Match      69.5%; Score 14.6; DB 8; Length 25;
Best Local Similarity 81.0%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
Db      21 TACATGCATGGGTTTCTGCC 1
```

```
RESULT 48
US-11-124-368A-8804/c
; Sequence 8804, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8804
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-8804
```

```
Query Match      69.5%; Score 14.6; DB 8; Length 201;
Best Local Similarity 81.0%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 TGCATGTATGGGTTATCTTCC 21
Db      187 TGAATGCATGGGTTTGTCTCC 167
```

```
RESULT 49
US-10-793-626-1891/c
; Sequence 1891, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
```

```

; SEQ ID NO 1891
; LENGTH: 866
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1891

```

Query Match	69.5%	Score 14.6;	DB 7;	Length 868;
Best Local Similarity	81.0%;	Pred. No. 3.3e+02;		
Matches 17:	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0;

QY 1 TGCATGTATGGGTATCTTCC 21  
DB 272 TGCACGTCTGAATTATCTTCC 252

RESULT 50  
US-10-793-626-1139/c  
; Sequence 1139, Application US/10793626  
: Publication No. US20050255478A1

```

1 GENERAL INFORMATION:
2 APPLICANT: KIMMERLY, WILLIAM JOHN
3 TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
4
5 FILE OF INVENTION: P034800S
6 CURRENT APPLICATION NUMBER: US/10/793,626
7 CURRENT FILING DATE: 2004-03-04
8 PRIOR APPLICATION NUMBER: 60/164,258
9 PRIOR FILING DATE: 1999-11-09
10 NUMBER OF SEQ ID NOS: 4472
11 SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 1139
; LENGTH: 912

```

; SERVOIII: 012  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence

```

; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
;
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-1139

```

Query Match 69.5%; Score 14.6; DB 7; Length 912;  
Best Local Similarity 81.0%; Pred. No. 3.4e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGCATGTATGGGTTATCTTCC 21  
||| | | | | | | | |  
nb 104 TGCACGCTGAATTTATCTTCC 84

RESULT 51  
US-10-793-626-2741/c  
; Sequence 2741, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION.

```

? GENERAL INFORMATION:
?
? APPLICANT: KIMMERLY, WILLIAM JOHN
?
? TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
?
? FILE REFERENCE: PU3480US
?
? CURRENT APPLICATION NUMBER: US/10/793,626
?
? CURRENT FILING DATE: 2004-03-04
?
? PRIOR APPLICATION NUMBER: 60/164,258
?
? PRIOR FILING DATE: 1999-11-09
?
? NUMBER OF SEQ ID NOS: 4472
?
? SOFTWARE: PatentIn ver. 2.1
?
? SEQ ID NO 2741
?
? LENGTH: 1080
?

```

```

; LENGTH: 1000
; TYPE: DNA
: ORGANISM: Artificial Sequence

```

; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
 ; OTHER INFORMATION: nucleic acid sequence  
 US-10-793-626-2741

US-10-793-626-2741

Query Match 69.5%; Score 14.6; DB 7; Length 1080;  
Best Local Similarity 81.0%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0

Qy 1 TGCATGTATGGGTATCTTCC 21  
272 TGCACGTCTGAATTATCTTCC 252

RESULT 52

US-10-750-185-62709/c  
; Sequence 62709, Application US/10750185  
; Publication No. US20050260603A1

```

, GENERAL INFORMATION:
, APPLICANT: MMI GENOMICS, INC.
, APPLICANT: DENISE, Sue K.
, APPLICANT: KERR, Richard
, APPLICANT: ROSENFELD, David
, APPLICANT: HOLM, Tom
, APPLICANT: BATES, Stephen
, APPLICANT: FANTIN, Dennis
, TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
, FILE REFERENCE: MM1100-2
, CURRENT APPLICATION NUMBER: US/10/750,185
, CURRENT FILING DATE: 2003-12-31

```

; CURRENT FILING DATE: 2003-12-31  
 ; PRIOR APPLICATION NUMBER: US 6  
 ; PRIOR FILING DATE: 2002-12-31

; PRIOR FILING DATE: 2002-12-31  
 ; NUMBER OF SEQ ID NOS: 64922  
 ; SOFTWARE: PatentIn version 3.1

```

; SOFTWARE: PATENTIN VERSION 3.1
; SEQ ID NO 62709
; LENGTH: 1145

```

```
;
; LENGTH: 1145
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
```

US-10-750-185-62709  
ORGANISM: Bovine  
19866880895343

Query Match	69.5%;	Score 14.6;	DB 7;	Length 1145;
Best Local Similarity	81.0%;	Pred. NO. 3.5e+02;		
Matches 17:	Conservative	0;	Mismatches 4;	Indels 0

QY  
Db

1 TGCATGATCGGTATCTTCC 21  
| | | | |  
41 TACATGATCGGTCTCTCCC 21

RESIST 53

RESULT 53  
US-10-750-623-62709/c  
; Sequence 62709, Application US/10750623  
; Publication No. US2005028751A1

; PUBLICATION NO: US20050267331  
 ; GENERAL INFORMATION:  
 : APPLICANT: MMI GENOMICS INC

APPLICANT: MMI GENOMICS, INC.  
APPLICANT: DENISE, Sue K.  
APPLICANT: KERR, Richard

APPLICANT: KERR, Richard  
APPLICANT: ROSENFELD, David  
APPLICANT: HOIM, Tom

; APPLICANT: HOLM, Tom  
 ; APPLICANT: BATES, Stephen

; APPLICANT: FANTIN, Dennis  
 ; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
 ; FILE REFERENCE: MM1100-1

; CURRENT APPLICATION NUMBER: US/10/750,623  
 ; CURRENT FILING DATE: 2003-12-31

; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31

; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1

```

; SEQ ID NO 62709
; LENGTH: 1145

```

TYPE: DNA  
ORGANISM: Bovine 19866880895343

US-10-750-623-62709

Query Match 69.5%; Score 14.6; DB 7; Length 1145;  
Best Local Similarity 81.0%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0

Matches	17: Conservative	0: Mismatches	4: Indels	0: Gaps
0: 0	0: 0	0: 0	0: 0	0: 0

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 41 TACATGTATGGGTCTCTCC 21

RESULT 54  
US-10-750-185-62508  
; Sequence 62508, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; PRIOR FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62508  
; LENGTH: 1208  
; TYPE: DNA  
; ORGANISM: Bovine 19866880913049  
US-10-750-185-62508

Query Match 69.5%; Score 14.6; DB 7; Length 1208;  
Best Local Similarity 81.0%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 1037 TGCATTCATGGGCTTCTTCC 1057

RESULT 55  
US-10-750-623-62508  
; Sequence 62508, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 62508  
; LENGTH: 1208  
; TYPE: DNA  
; ORGANISM: Bovine 19866880913049  
US-10-750-623-62508

Query Match 69.5%; Score 14.6; DB 7; Length 1208;  
Best Local Similarity 81.0%; Pred. No. 3.5e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 1037 TGCATTCATGGGCTTCTTCC 1057

Db 1037 TGCATTCATGGGCTTCTTCC 1057

RESULT 56  
US-10-750-185-29432  
; Sequence 29432, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 29432  
; LENGTH: 1275  
; TYPE: DNA  
; ORGANISM: Bovine 19866881425942  
US-10-750-185-29432

Query Match 69.5%; Score 14.6; DB 7; Length 1275;  
Best Local Similarity 81.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 997 TGAATGTATGGGTTTTTTTC 1017

RESULT 57  
US-10-750-623-29432  
; Sequence 29432, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 29432  
; LENGTH: 1275  
; TYPE: DNA  
; ORGANISM: Bovine 19866881425942  
US-10-750-623-29432

Query Match 69.5%; Score 14.6; DB 7; Length 1275;  
Best Local Similarity 81.0%; Pred. No. 3.6e+02;  
Matches 17; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGCATGTATGGGTATCTTCC 21  
Db 997 TGAATGTATGGGTTTTTTTC 1017

RESULT 58

```

US-10-750-185-62216
; Sequence 62216, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERSE, Richard
; APPLICANT: ROSENFIELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFECTION
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62216
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-62216

```

Query Match 69.5%; Score 14.6; DB 7; Length 1353;  
Best Local Similarity 81.0%; Pred. No. 3.6e+02;  
Matches 17: Conservative 0; Mismatches 4; Indels 0

QY 1 TGCATGTATGGGTATCTTCC 21  
299 TGCTTGAATGCCTTATCTTCC 319  
Dp

## RESULT 59

```

US-10-750-623-62216
; Sequence 62216, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: Denise, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 62216
; LENGTH: 1353
; TYPE: DNA
; ORGANISM: Bovine 19866880904068
US-10-750-623-62216

```

Query Match	69.5%	Score 14.6;	DB 7;	Length 1353;
Best Local Similarity	81.0%;	Pred. No. 3.6e+02;		
Matches 17: Conservative	0;	Mismatches 4;	Indels 0;	Gaps 0;

**Qy**

1 TGCATGTATGGGTTATCTTCC 21  
||| ||| ||| ||| ||| ||| |||

**D<sub>b</sub>**

299 TGCTTGAATGCCATTATCTTCC 319

## RESULT 60

US-10-512-109-5/c  
; Sequence 5, Application US/10512109

**This Page Blank (uspto)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:49:12 ; Search time 2026.12 Seconds  
(without alignments)  
577.298 Million cell updates/sec

Title: US-10-716-005-3  
Perfect score: 25  
Sequence: 1 caaattaaagagactattcgtaa 25

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_est3: \*  
4: gb\_est4: \*  
5: gb\_est5: \*  
6: gb\_est6: \*  
7: gb\_est7: \*  
8: gb\_est8: \*  
9: gb\_est9: \*  
10: gb\_gss2: \*  
11: gb\_gss3: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.8	79.2	281	10	CZ594841 CMHD-GT 2
C 2	19.8	79.2	388	10	CZ317624 ZMMBF0017
C 3	19.2	76.8	394	10	CG838808 ZMMBSc025
C 4	19.2	76.8	399	10	CL453893 ZMMBSc048
C 5	19.2	76.8	473	9	AZ932227 474. dbz96
C 6	19.2	76.8	565	10	CZ687653 OA_Bba016
C 7	19.2	76.8	582	3	BM331796 MEST166-G
C 8	19.2	76.8	600	3	BM341275 MEST332-G
C 9	19.2	76.8	612	11	CR340097 mte1-71J2
C 10	19.2	76.8	618	3	BM349126 MEST308-G
C 11	19.2	76.8	642	11	DE026109 Branchios
C 12	19.2	76.8	673	3	BM348442 MEST230-H
C 13	19.2	76.8	765	10	CL680700 PRI012a.D
C 14	19.2	76.8	785	10	CNS02PKY Tetraodon
C 15	19.2	76.8	804	11	CR146965 Forward s
C 16	19.2	76.8	810	9	BZ784243 PUCAT94TB
C 17	19.2	76.8	812	11	CR173895 Forward s
C 18	19.2	76.8	815	10	CG348543 OG0FUL3TV
C 19	19.2	76.8	816	10	CG348533 OG0FUL3TH
C 20	19.2	76.8	872	9	CG681145 OGUFK42TV
C 21	19.2	76.8	895	2	BG479888 602527434
C 22	19.2	76.8	925	9	CG681138 OGUFK42TH

C 23	19.2	76.8	991	9	CC686212	CC686212
C 24	18.8	75.2	444	10	CG033706	CG033706
C 25	18.8	75.2	460	9	AQ989224	AQ989224
C 26	18.8	75.2	612	9	AQ233842	AQ233842
C 27	18.6	74.4	148	7	CV709364	CV709364
C 28	18.6	74.4	206	9	AZ794021	AZ794021
C 29	18.6	74.4	537	11	D8043516	D8043516
C 30	18.6	74.4	569	9	BZ615728	BZ615728
C 31	18.6	74.4	586	6	CB092313	CB092313
C 32	18.6	74.4	600	8	CX543634	CX543634
C 33	18.6	74.4	604	8	CX644973	CX644973
C 34	18.6	74.4	608	8	CX644972	CX644972
C 35	18.6	74.4	666	7	CJ340874	CJ340874
C 36	18.6	74.4	670	3	BI390310	BI390310
C 37	18.6	74.4	670	9	CC373687	CC373687
C 38	18.6	74.4	691	1	AW333363	AW333363
C 39	18.6	74.4	706	8	CX546303	CX546303
C 40	18.6	74.4	708	7	CJ397741	CJ397741
C 41	18.6	74.4	712	7	CV509866	CV509866
C 42	18.6	74.4	712	10	CZ287609	CZ287609
C 43	18.6	74.4	717	7	CV706617	CV706617
C 44	18.6	74.4	727	7	CV709207	CV709207
C 45	18.6	74.4	731	7	CJ344248	CJ344248
C 46	18.6	74.4	732	7	CJ357933	CJ357933
C 47	18.6	74.4	733	8	CX641978	CX641978
C 48	18.6	74.4	740	10	CZ279576	CZ279576
C 49	18.6	74.4	742	7	CJ361428	CJ361428
C 50	18.6	74.4	745	5	BUI330653	BUI330653
C 51	18.6	74.4	754	7	CJ358434	CJ358434
C 52	18.6	74.4	758	9	BZ246552	BZ246552
C 53	18.6	74.4	764	7	CV708436	CV708436
C 54	18.6	74.4	788	8	CX644101	CX644101
C 55	18.6	74.4	793	7	CV706539	CV706539
C 56	18.6	74.4	798	7	CV705522	CV705522
C 57	18.6	74.4	798	9	BZ536243	BZ536243
C 58	18.6	74.4	810	7	CV704979	CV704979
C 59	18.6	74.4	816	7	CV708661	CV708661
C 60	18.6	74.4	823	7	CJ401436	CJ401436

#### ALIGNMENTS

#### RESULT 1

CZ594841/c  
LOCUS CMHD-GT-220D12-3 281 bp mRNA linear GSS 30-JUN-2005  
DEFINITION CMHD-GT-220D12-3 3', mRNA sequence.

#### ACCESSION

VERSION CZ594841.1 GI:68367780

#### KEYWORDS

GSS.

#### SOURCE

Mus musculus (house mouse)

#### ORGANISM

Mus musculus

#### REFERENCE

1 (bases 1 to 281)

#### AUTHORS

Stanford, W.L.

#### TITLE

Unpublished (2002)

#### JOURNAL

Contact: Stanford WL

#### COMMENT

Institute of Biomaterials & Biomedical Engineering

University of Toronto

407 Rosebrugh Bldg., 4 Taddle Creek Rd., Toronto, Ontario, Canada

M5S 3G9

Tel: 416 946 8379

Fax: 416 978 4317

Email: william.stanford@utoronto.ca

pcgTNDf Gene trap insertion. The sequence tag is generated by 3'

race. The ES cell line harboring this insertion of the target gene

is available through the following web site:

http://pokey.ihme.utoronto.ca/sequence\_report.php?id=220D12.

Class: Gene Trap.

FEATURES source	Location/Qualifiers		Location/Qualifiers	
	1..281		394 bp DNA linear GSS 12-NOV-2003	
	/organism="Mus musculus"		ZMMBbC0251K21r ZMMBbC (EcoRI) Zea mays genomic clone ZMMBbC0251K21	
	/mol_type="mRNA"		3', genomic survey sequence.	
	/strain="129S3"		CG838808	
	/db_xref="taxon:10090"		CG838808	
	/clone="CMHD-GT_220D12-3"		CG838808.1 GI:38303207	
	/sex="Male"		GSS.	
	/cell_type="Embryonic Stem Cell"		Zea mays	
	/cell_line="R1"		Zea mays	
ORIGIN	/clone_lib="GTL_R1_pGTMDF"		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	/note="Vector: pGTMDF"		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD	
	Query Match 79.2%; Score 19.8; DB 10; Length 281;		clade; Panicoideae; Andropogoneae; Zea.	
	Best Local Similarity 91.3%; Pred. No. 3e+02;		1 (bases 1 to 394)	
	Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,	
	Qy 2 AAATTAAGAGACTATTTCGTGCA 24		Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.	
			Sequencing of the maize genome at PGIR (2003c)	
	Db 48 AAATTAAGAGACTATTTCGTGCA 26		Unpublished (2003)	
			Contact: Bharti,A.K.	
			Dr.Joachim Messing's lab	
FEATURES source	Location/Qualifiers		The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers	
	1..388		University	
	/organism="Zea mays"		190 Frelinghuysen Road, Piscataway, NJ 08854, USA	
	/mol_type="genomic DNA"		Tel: 732 445 3801	
	/cultivar="B73"		Fax: 732 445 5735	
	/db_xref="taxon:4577"		Email: bharti@waksman.rutgers.edu	
	/clone="ZMMBbC0251K21"		Seq primer: SP6	
	/lab_host="E. coli DH10B"		Class: BAC ends	
	/clone_lib="ZMMBbC (EcoRI)"		High quality sequence start: 118.	
	/note="Vector: pTABAC2.1; Site_1: EcoRI; Site_2: EcoRI"		Location/Qualifiers	
ORIGIN	1..394		1..394	
	Query Match 76.8%; Score 19.2; DB 10; Length 394;		/organism="Zea mays"	
	Best Local Similarity 87.5%; Pred. No. 5.8e+02;		/mol_type="genomic DNA"	
	Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;		/cultivar="B73"	
	Qy 2 AAATTAAGAGACTATTTCGTGCA 25		/db_xref="taxon:4577"	
			/clone="ZMMBbC0251K21"	
	Db 131 AAATTAAGAGACTATTTCGTGCA 108		/lab_host="E. coli DH10B"	
			/clone_lib="ZMMBbC (EcoRI)"	
			/note="Vector: pTABAC2.1; Site_1: EcoRI; Site_2: EcoRI"	
FEATURES source	Location/Qualifiers		Location/Qualifiers	
	1..388		399 bp DNA linear GSS 29-MAR-2004	
	/organism="Zea mays"		ZMMBbB0485M20f ZMMBbB (HindIII) Zea mays genomic clone	
	/mol_type="genomic DNA"		ZMMBbB0485M20 5', genomic survey sequence.	
	/cultivar="B73"		CL453893	
	/db_xref="taxon:4577"		CL453893	
	/clone="ZMMBf0017D18"		CL453893.1 GI:45780244	
	/lab_host="EPI100-T1"		GSS.	
	/clone_lib="ZMMBf"		Zea mays	
	/note="Vector: pEpiFOS-5; Site_1: Eco72I"		Zea mays	
ORIGIN	1..388		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
	Query Match 79.2%; Score 19.8; DB 10; Length 388;		Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD	
	Best Local Similarity 91.3%; Pred. No. 3.1e+02;		clade; Panicoideae; Andropogoneae; Zea.	
	Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		1 (bases 1 to 399)	
	Qy 2 AAATTAAGAGACTATTTCGTGCA 24		Bharti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,	
			Zohovetz,V., Fuks,G., Yu,Y., Wing,R. and Messing,J.	
	Db 204 AAATTAAGAGACAAATTCATGCA 226		Sequencing of the maize genome at PGIR (2003c)	
			Unpublished (2003)	
			Contact: Bharti,A.K.	
			Dr.Joachim Messing's lab	
			The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers	
			University	
			190 Frelinghuysen Road, Piscataway, NJ 08854, USA	

Tel: 732 445 3801  
 Fax: 732 445 5735  
 Email: bharti@wakeman.rutgers.edu  
 Seq primer: T7  
 Class: BAC ends  
 High quality sequence start: 104.  
 Location/Qualifiers

## FEATURES

## source

1..399  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="ZMMBBb0485M20"  
 /lab\_host="E. coli DH10B"  
 /clone\_lib="ZMMBBb (HindIII)"  
 /notes="Vector: pCUGI; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 399;  
 Best Local Similarity 87.5%; Pred. No. 5.8e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 212 CAAATTAAGAGACTATTCGTGCA 189

## RESULT 5

## LOCUS

DEFINITION 474.dhz96c07.s1 Saccharomyces unisporus NRRL Y-1556 Saccharomyces unisporus genomic clone 474.dhz96c07.s1, genomic survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Saccharomyces unisporus  
 Saccharomyces unisporus  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

## REFERENCE

## AUTHORS

1 (bases 1 to 473)  
 Clifton, P.F., Hillier, L.W., Fulton, L., Graves, T., Miner, T.,  
 Gish, W.R., Waterston, R.H. and Johnston, M.  
 Surveying Saccharomyces genomes to identify functional elements by  
 comparative DNA sequence analysis

## TITLE

## JOURNAL

## COMMENT

Unpublished (2001)  
 Contact: Johnston M  
 Department of Genetics  
 Washington University Medical School  
 Box 8232, 4566 Scott Ave., St. Louis, MO 63110, USA  
 Tel: 314 362 2735  
 Fax: 314 362 7855  
 Email: mj@genetics.wustl.edu  
 Class: random plasmid subclone.  
 Location/Qualifiers

## FEATURES

## source

1..473  
 /organism="Saccharomyces unisporus"  
 /mol\_type="genomic DNA"  
 /strain="NRRL Y-1556 (CBS 398)"  
 /db\_xref="taxon:27294"  
 /clone="474.dhz96c07.s1"  
 /clone\_lib="Saccharomyces unisporus NRRL Y-1556"  
 /notes="Random genomic sequence"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 473;  
 Best Local Similarity 87.5%; Pred. No. 5.8e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25

||||| ||||| ||||| ||||| |||||

Db 335 AAATTAAGAGATTATACATGCAA 358

## RESULT 6

## LOCUS

## DEFINITION

OA\_BBa0164M21.r OA\_BBa Oryza alta genomic clone OA\_BBa0164M21 3',  
 genomic survey sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Oryza alta  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhacoidae; Oryzae; Oryza.

1 (bases 1 to 565)  
 SanMiguel, P., Westerman, R., Kim, H., Yu, Y., Wissotski, M., Yost, D.,  
 Stum, D., Rao, K., Luo, M., Jetty, R., Kudrna, D., Muller, C.,  
 Hatfield, J., Soderlund, C., Wing, R. and Jackson, S.A.  
 OMAP (Oryza Map Alignment Project)- Purdue University

Unpublished (2004)  
 Contact: Scott A. Jackson  
 Jackson Laboratory  
 Purdue University  
 915 W. State St., West Lafayette, IN 47907, USA  
 Tel: 7654963621  
 Fax: 7654967255  
 Email: sjackson@purdue.edu  
 Basecalling by phred version 0.020425.c. This sequence was derived  
 from the raw sequence read by clipping with Lucy version 1.19a.  
 Bases 114-678 of the raw sequence (length 1100) were retained after  
 clipping.  
 Plate: 0164 row: M column: 21  
 Seq primer: CAC TCA TTA GGC ACC CCA  
 Class: BAC ends.  
 Location/Qualifiers

1..565  
 /organism="Oryza alta"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:52545"  
 /clone="OA\_BBa0164M21"  
 /tissue\_type="young leaves"  
 /lab\_host="DH10B-T1 phage resistant"  
 /clone\_lib="OA\_BBa"  
 /note="Vector: pAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## FEATURES

## source

1..565  
 /organism="Oryza alta"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:52545"  
 /clone="OA\_BBa0164M21"  
 /tissue\_type="young leaves"  
 /lab\_host="DH10B-T1 phage resistant"  
 /clone\_lib="OA\_BBa"  
 /note="Vector: pAGIBAC1; Site\_1: HindIII; Site\_2: HindIII"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 565;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25

||||| ||||| ||||| ||||| |||||

Db 445 AAATTAAGAGATTATTCATGCCA 422

## RESULT 7

## LOCUS

## DEFINITION

BM331796  
 MEST166-G03.T3 ISUM5-RN Zea mays cDNA clone MEST166-G03 3', mRNA  
 sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
 clade; Panicoidae; Andropogoneae; Zea.

1 (bases 1 to 582)  
 Wen, T.J., Qiu, F., Guo, L., Ashlock, D.A. and Schnable, P.S.  
 Expressed Sequence Tags from B73 Maize: various stages and tissues  
 including seedlings treated with a variety of hormones  
 Unpublished (2001)  
 Contact: Patrick S. Schnable

BM331796  
 BM331796  
 EST.  
 Zea mays  
 Zea mays

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 565;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25

||||| ||||| ||||| ||||| |||||

Db 445 AAATTAAGAGATTATTCATGCCA 422

Schnable Laboratory  
Iowa State University  
2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA  
Tel: 515-294-0975  
Fax: 515-294-5256  
Email: schnable@iastate.edu  
Individual basecall and confidence value were assigned using the Phred software,  
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>)#b  
rt). Overall sequence quality assessment and vector trimming were conducted using the Lucy software (<http://www.tigr.org/softlab/>). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high-quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers.  
PCR Primers  
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)  
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

#### FEATURES

source  
1..582  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="MEST166-G03"  
/tissue\_type="mixed"  
/lab\_host="DH10B"  
/clone\_lib="ISUM5-RN"  
/notes="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI;  
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),  
Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels  
(3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65  
DAG), Tassel (3-39 cm, 53 and 56 DAG), Immature ear  
(0.2-3.0 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk,  
unpollinated first ear, ear shank, etiolated seedlings,  
callus, Cycloheximide-treated callus, Anaerobic treated  
seedlings, NAA (a-Naphthalene acetic acid)-treated  
seedlings, Kinetin-treated seedlings, ACP  
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,  
Brassinolide-treated seedlings, ABA (Abscissic  
acid)-treated seedlings, GA (Gibberellic acid)-treated  
seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA  
molecules were generated as follows. First-strand cDNA was  
prepared from oligo-dT selected mRNA by priming with a  
NotI oligo-dT primer (5'  
AACTGGAAGAAATCGCGCCGAGGAATTTTTTTTTTTT). The  
resulting DNA:RNA hybrid was treated with RNase H and used  
as a template for DNA PolI-catalyzed second strand  
synthesis. After the addition of EcoRI adaptors, the  
ds-cDNAs were digested with NotI and size-selected. The  
resulting molecules were directionally cloned into the  
EcoRI and NotI sites of the pT7T3PAC vector. The library  
then went through one round of normalization to Cot value  
of 5 based on the methods of Marcelo Bento Soares (Genome  
Research 6: 791-806, 1996)."

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 582;  
Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 1 CAAATTAAGAGACTATTCGTGCA 24  
||||||| ||||| ||||| |||||  
Db 128 CAAATTAAGACTATGCGTGCA 151

#### RESULT 8

BM341275 600 bp mRNA linear EST 16-JAN-2002  
LOCUS MEST332-G07.T3 ISUM5-RN Zea mays cDNA clone MEST332-G07 3', mRNA  
DEFINITION sequence.

#### ACCESSION VERSION KEYWORDS SOURCE ORGANISM

BM341275  
BM341275.1 GI:18171435  
EST.  
Zea mays  
Zea mays

#### REFERENCE

Wen, T.J., Qiu, F., Guo, L., Ashlock, D.A. and Schnable, P.S.  
Expressed Sequence Tags from B73 Maize: various stages and tissues  
including seedlings treated with a variety of hormones  
Unpublished (2001)

#### JOURNAL

Contact: Patrick S. Schnable

#### COMMENT

Schnable Laboratory  
Iowa State University  
2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA  
Tel: 515-294-0975  
Fax: 515-294-5256  
Email: schnable@iastate.edu

Individual basecall and confidence value were assigned using the Phred software,  
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>)#b  
rt). Overall sequence quality assessment and vector trimming were conducted using the Lucy software (<http://www.tigr.org/softlab/>). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high-quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers.

PCR Primers  
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)  
BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).  
Location/Qualifiers  
1..600  
/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="MEST332-G07"  
/tissue\_type="mixed"  
/lab\_host="DH10B"  
/clone\_lib="ISUM5-RN"  
/notes="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI;  
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),  
Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels  
(3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65  
DAG), Tassel (3-39 cm, 53 and 56 DAG), Immature ear  
(0.2-3.0 cm, 53, 56, 59 DAG), Husk (73 DAG), Silk,  
unpollinated first ear, ear shank, etiolated seedlings,  
callus, Cycloheximide-treated callus, Anaerobic treated  
seedlings, NAA (a-Naphthalene acetic acid)-treated  
seedlings, Kinetin-treated seedlings, ACP  
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,  
Brassinolide-treated seedlings, ABA (Abscissic  
acid)-treated seedlings, GA (Gibberellic acid)-treated  
seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA  
molecules were generated as follows. First-strand cDNA was  
prepared from oligo-dT selected mRNA by priming with a  
NotI oligo-dT primer (5'  
AACTGGAAGAAATCGCGCCGAGGAATTTTTTTTTTTT). The  
resulting DNA:RNA hybrid was treated with RNase H and used  
as a template for DNA PolI-catalyzed second strand  
synthesis. After the addition of EcoRI adaptors, the  
ds-cDNAs were digested with NotI and size-selected. The  
resulting molecules were directionally cloned into the  
EcoRI and NotI sites of the pT7T3PAC vector. The library  
then went through one round of normalization to Cot value  
of 5 based on the methods of Marcelo Bento Soares (Genome  
Research 6: 791-806, 1996)."

#### FEATURES

source

#### ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 600;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 127 CAAATTAAATGGACTATGCGTGCA 150

RESULT 9  
 CR340097/c  
 LOCUS  
 DEFINITION mte1-71J2RM1 BAC end, cultivar Jemalong A17 of Medicago truncatula,  
 genomic survey sequence.

ACCESSION CR340097

VERSION CR340097.1 GI:44910432

KEYWORDS GSS.

SOURCE Medicago truncatula (barrel medic)

ORGANISM  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE 1 (bases 1 to 612)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (25-FEB-2004) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)

FEATURES

source

1..612  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="Jemalong A17"  
 /db\_xref="taxon:3880"  
 /clone\_lib="MTEI"  
 /notes="Vector: pindigobac ; Site\_1: EcoRI ; Site\_2: EcoRI  
 ; Debelle F. and Chalhou B.  
 Genoscope sequence ID : mte1-71J2RM1"

ORIGIN

Query Match 76.8%; Score 19.2; DB 11; Length 612;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 355 CAAATTAAATATCTATTCTGTGCA 332

RESULT 10  
 BM349126  
 LOCUS  
 DEFINITION BM3T308-G10.T3 ISUM5-RN Zea mays cDNA clone MEST308-G10 3', mRNA  
 sequence.

ACCESSION BM349126

VERSION BM349126.1 GI:18173738

KEYWORDS EST.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 618)

AUTHORS Wen,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.

TITLE Expressed Sequence Tags from B73 Maize: various stages and tissues  
 including seedlings treated with a variety of hormones

JOURNAL Unpublished (2001)

COMMENT Contact: Patrick S. Schnable

Schnable Laboratory

Iowa State University

2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA

Tel: 515-294-0975

Fax: 515-294-5256  
 Email: schnable@iastate.edu  
 Individual basecall and confidence value were assigned using the  
 Phred software,  
 (<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#b  
 rt). Overall sequence quality assessment and vector trimming were  
 conducted using the Lucy software (<http://www.tigr.org/softlab/>).  
 Lucy parameters were set to ensure an overall trimmed quality of  
 97.5% or better without any vector fragments in the chosen  
 high-quality region of each sequence. Low-quality bases between the  
 poly-T and the high-quality region were replaced with N's to serve  
 as spacers.

PCR Primers  
 FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)  
 BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
 Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

Location/Qualifiers  
 1..618  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="MEST308-G10"  
 /tissue\_type="mixed"  
 /lab\_host="DH108"  
 /clone\_lib="ISUM5-RN"  
 /note="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI;  
 Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),  
 Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels  
 (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65  
 DAG), Tassel (3-39 cm, 53 and 56 DAG), Husk (73 DAG), Silk,  
 (0.2-3.0 cm, 53, 56, 59 DAG), unpollinated first ear, etiolated seedlings,  
 callus, Cycloheximide-treated callus, Anaerobic treated  
 seedlings, NAA (a-Naphthalene acetic acid)-treated  
 seedlings, Kinetin-treated seedlings, ACPG  
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,  
 Brassinolide-treated seedlings, ABA (Abscissic  
 acid)-treated seedlings, GA (Gibberellic acid)-treated  
 seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA  
 molecules were generated as follows. First-strand cDNA was  
 prepared from oligo-dT selected mRNA by priming with a  
 NotI oligo-dT primer (5'  
 AACTGGAAGATTCGCGCCGCGAGGAATTTTTTTTTTTTTTTT). The  
 resulting DNA:RNA hybrid was treated with RNase H and used  
 as a template for DNA PolI-catalyzed second strand  
 synthesis. After the addition of EcoRI adaptors, the  
 ds-cDNAs were digested with NotI and size-selected. The  
 resulting molecules were directionally cloned into the  
 EcoRI and NotI sites of the pT7T3PAC vector. The library  
 then went through one round of normalization to Cot value  
 of 5 based on the methods of Marcelo Bento Soares (Genome  
 Research 6: 791-806, 1996)."

Research 6: 791-806, 1996)."  
 Query Match 76.8%; Score 19.2; DB 3; Length 618;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

FEATURES

source

1..618  
 /organism="Zea mays"  
 /mol\_type="mRNA"  
 /cultivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="MEST308-G10"  
 /tissue\_type="mixed"  
 /lab\_host="DH108"  
 /clone\_lib="ISUM5-RN"  
 /note="Vector: pT7T3PAC; Site\_1: EcoRI; Site\_2: NotI;  
 Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),  
 Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels  
 (3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65  
 DAG), Tassel (3-39 cm, 53 and 56 DAG), Husk (73 DAG), Silk,  
 (0.2-3.0 cm, 53, 56, 59 DAG), unpollinated first ear, etiolated seedlings,  
 callus, Cycloheximide-treated callus, Anaerobic treated  
 seedlings, NAA (a-Naphthalene acetic acid)-treated  
 seedlings, Kinetin-treated seedlings, ACPG  
 (1-aminocyclopropane-1-carboxylic acid)-treated seedlings,  
 Brassinolide-treated seedlings, ABA (Abscissic  
 acid)-treated seedlings, GA (Gibberellic acid)-treated  
 seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA  
 molecules were generated as follows. First-strand cDNA was  
 prepared from oligo-dT selected mRNA by priming with a  
 NotI oligo-dT primer (5'  
 AACTGGAAGATTCGCGCCGCGAGGAATTTTTTTTTTTTTTTT). The  
 resulting DNA:RNA hybrid was treated with RNase H and used  
 as a template for DNA PolI-catalyzed second strand  
 synthesis. After the addition of EcoRI adaptors, the  
 ds-cDNAs were digested with NotI and size-selected. The  
 resulting molecules were directionally cloned into the  
 EcoRI and NotI sites of the pT7T3PAC vector. The library  
 then went through one round of normalization to Cot value  
 of 5 based on the methods of Marcelo Bento Soares (Genome  
 Research 6: 791-806, 1996)."

ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 618;  
 Best Local Similarity 87.5%; Pred. No. 5.9e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTGCA 24  
 ||||| ||||| ||||| ||||| |||||  
 Db 124 CAAATTAAATGGACTATGCGTGCA 147

RESULT 11

DE026109

LOCUS

DEFINITION

DE026109

Branchiostoma floridae DNA, clone: CH302-041F16.F, genomic survey

sequence.

DE026109

VERSION

DE026109.1 GI:62269579

KEYWORDS GSS.

SOURCE Branchiostoma floridae (Florida lancelet)

```

ORGANISM      Branchiostoma floridae
               Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
               Branchiostoma.
REFERENCE
AUTHORS      Fujiyama,A., Toyoda,A., Hattori,M. and Sakaki,Y.
TITLE        BAC end sequences of CHORI-302 Amphioxus Library
JOURNAL      Published Only in Database (2005)
REFERENCE
AUTHORS      Fujiyama,A.
TITLE        Direct Submission
JOURNAL      Submitted (31-MAR-2005) Asao Fujiyama, The Institute of Physical
               and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
               1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan
               (E-mail:afujiyamegsc.riken.jp URL:http://hgp.sec.riken.jp/,
               Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      PRIMERS
               Sequencing : T7
LIBRARY
Vector       : pTARBAC2.1
R.Site 1    : ECORI
R.Site 2    : ECORI.
Location/Qualifiers
1. .642
/organism="Branchiostoma floridae"
/mol_type="genomic DNA"
/db_xref="taxon:7739"
/clone="CH302-041F16.F"
/sex="male"
/tissue_type="sperm"
/clone_lib="CHORI0302 Amphioxus genomic BAC library"
/note="common name:amphioxus"
ORIGIN
Query Match 76.8%; Score 19.2; DB 11; Length 642;
Best Local Similarity 87.5%; Pred. No. 5.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAATTAAAGAGACTATTCTGTGCA 25
Db 211 AAGTTAAAGAAACTATTCTGTGCA 234

RESULT 12
BM348442
LOCUS      MEST290-H12.T3 ISUM5-RN Zea mays cDNA clone MEST290-H12 3', mRNA
DEFINITION
sequence.
ACCESSION  BM348442
VERSION    BM348442.1 GI:18173054
KEYWORDS   EST.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 673)
Wen,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.
Expressed Sequence Tags from B73 Maize: various stages and tissues
including seedlings treated with a variety of hormones
Unpublished (2001)
Contact: Patrick S. Schnable
Schnable Laboratory
Iowa State University
2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA
Tel: 515-294-0975
Fax: 515-294-5256
Email: schnable@iastate.edu
Individual basecall and confidence value were assigned using the
Phred software,
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#b
rt). Overall sequence quality assessment and vector trimming were
conducted using the Lucy software (<http://www.tigr.org/softlab/>).
Lucy parameters were set to ensure an overall trimmed quality of

ORGANISM      Branchiostoma floridae
               Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
               Branchiostoma.
REFERENCE
AUTHORS      Fujiyama,A., Toyoda,A., Hattori,M. and Sakaki,Y.
TITLE        BAC end sequences of CHORI-302 Amphioxus Library
JOURNAL      Published Only in Database (2005)
REFERENCE
AUTHORS      Fujiyama,A.
TITLE        Direct Submission
JOURNAL      Submitted (31-MAR-2005) Asao Fujiyama, The Institute of Physical
               and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
               1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan
               (E-mail:afujiyamegsc.riken.jp URL:http://hgp.sec.riken.jp/,
               Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      PRIMERS
               Sequencing : T7
LIBRARY
Vector       : pTARBAC2.1
R.Site 1    : ECORI
R.Site 2    : ECORI.
Location/Qualifiers
1. .642
/organism="Branchiostoma floridae"
/mol_type="genomic DNA"
/db_xref="taxon:7739"
/clone="CH302-041F16.F"
/sex="male"
/tissue_type="sperm"
/clone_lib="CHORI0302 Amphioxus genomic BAC library"
/note="common name:amphioxus"
ORIGIN
Query Match 76.8%; Score 19.2; DB 11; Length 642;
Best Local Similarity 87.5%; Pred. No. 5.9e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 AAATTAAAGAGACTATTCTGTGCA 25
Db 211 AAGTTAAAGAAACTATTCTGTGCA 234

RESULT 12
BM348442
LOCUS      MEST290-H12.T3 ISUM5-RN Zea mays cDNA clone MEST290-H12 3', mRNA
DEFINITION
sequence.
ACCESSION  BM348442
VERSION    BM348442.1 GI:18173054
KEYWORDS   EST.
SOURCE     Zea mays
ORGANISM   Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 673)
Wen,T.J., Qiu,F., Guo,L., Ashlock,D.A and Schnable,P.S.
Expressed Sequence Tags from B73 Maize: various stages and tissues
including seedlings treated with a variety of hormones
Unpublished (2001)
Contact: Patrick S. Schnable
Schnable Laboratory
Iowa State University
2035B R.J. Carver Co-Lab, Ames, IA 50011-3650, USA
Tel: 515-294-0975
Fax: 515-294-5256
Email: schnable@iastate.edu
Individual basecall and confidence value were assigned using the
Phred software,
(<http://depts.washington.edu/ventures/collabtr/direct/index.htm>#b
rt). Overall sequence quality assessment and vector trimming were
conducted using the Lucy software (<http://www.tigr.org/softlab/>).
Lucy parameters were set to ensure an overall trimmed quality of

FEATURES
source
1. .673
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MEST290-H12"
/tissue_type="mixed"
/lab_host="DH10B"
/clone_lib="ISUM5-RN"
/note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),
Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels
(3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65
DAG), Tassel (3-39 cm, 53 and 56 DAG), Husk (73 DAG), Silk,
(0.2-3.0 cm, 53, 56, 59 DAG), Husk (73 DAG), Immature ear
unpollinated first ear, ear shank, etiolated seedlings,
callus, Cycloheximide-treated callus, Anaerobic treated
seedlings, NAA (a-Naphthalene acetic acid)-treated
seedlings, Kinetin-treated seedlings, ACPG
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,
Brassinolide-treated seedlings, ABA (Abscissic
acid)-treated seedlings, GA (Gibberellic acid)-treated
seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA
molecules were generated as follows. First-strand cDNA was
prepared from oligo-dt selected mRNA by priming with a
NotI oligo-dT primer (5'
AATCGAAGATTCGCGCGCAGCAATTTTTTTTTTTTTT). The
resulting DNA:RNA hybrid was treated with RNase H and used
as a template for DNA Pol-I-catalyzed second strand
synthesis. After the addition of EcoRI adaptors, the
ds-cDNAs were digested with NotI and size-selected. The
resulting molecules were directionally cloned into the
EcoRI and NotI sites of the pT73PAC vector. The library
then went through one round of normalization to Cot value
of 5 based on the methods of Marcelo Bento Soares (Genome
Research 6: 791-806, 1996)."
ORIGIN
Query Match 76.8%; Score 19.2; DB 3; Length 673;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 CAATTAAAGAGACTATTCTGTGCA 24
Db 129 CAATTAAAGAGACTATTCTGTGCA 152

RESULT 13
CL680700
LOCUS      PRI012a.D11.2 - PRI012a.BR (765) Mixed stage fosmid library of P.
DEFINITION
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION  CL680700
VERSION    CL680700.1 GI:50187660
KEYWORDS   GSS.
SOURCE     Pristionchus pacificus
ORGANISM   Pristionchus pacificus
            Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
            Neodiplogasteridae; Pristionchus.
1 (bases 1 to 765)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AppADB: an AcedB database for the nematode satellite organism
Pristionchus pacificus

```

97.5% or better without any vector fragments in the chosen high-quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers.

PCR Primers  
 FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAA TAG)  
 BACKWARD: primer T3 (ATT AAC CCT CAC TAA AG)  
 Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

Location/Qualifiers  
 1. .673

/organism="Zea mays"  
/mol\_type="mRNA"  
/cultivar="B73"  
/db\_xref="taxon:4577"  
/clone="MEST290-H12"  
/tissue\_type="mixed"  
/lab\_host="DH10B"  
/clone\_lib="ISUM5-RN"

/note="Vector: pT73PAC; Site 1: EcoRI; Site 2: NotI;  
Tissues: Germinated seed and seedlings (1, 2, 8, 11 DAG),  
Mixed mature tissues (17, 21, 38, 69, 77 DAG), Kernels  
(3, 5, 10, 15, 20, 25, 30, DAP), Adventitious roots (65  
DAG), Tassel (3-39 cm, 53 and 56 DAG), Husk (73 DAG), Silk,  
(0.2-3.0 cm, 53, 56, 59 DAG), Husk (73 DAG), Immature ear  
unpollinated first ear, ear shank, etiolated seedlings,  
callus, Cycloheximide-treated callus, Anaerobic treated  
seedlings, NAA (a-Naphthalene acetic acid)-treated  
seedlings, Kinetin-treated seedlings, ACPG  
(1-aminocyclopropane-1-carboxylic acid)-treated seedlings,  
Brassinolide-treated seedlings, ABA (Abscissic  
acid)-treated seedlings, GA (Gibberellic acid)-treated  
seedlings, JA (Jasmonic acid)-treated seedlings. ds-cDNA  
molecules were generated as follows. First-strand cDNA was  
prepared from oligo-dt selected mRNA by priming with a  
NotI oligo-dT primer (5'  
AATCGAAGATTCGCGCGCAGCAATTTTTTTTTTTTTT). The  
resulting DNA:RNA hybrid was treated with RNase H and used  
as a template for DNA Pol-I-catalyzed second strand  
synthesis. After the addition of EcoRI adaptors, the  
ds-cDNAs were digested with NotI and size-selected. The  
resulting molecules were directionally cloned into the  
EcoRI and NotI sites of the pT73PAC vector. The library  
then went through one round of normalization to Cot value  
of 5 based on the methods of Marcelo Bento Soares (Genome  
Research 6: 791-806, 1996)."

ORIGIN

Query Match 76.8%; Score 19.2; DB 3; Length 673;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAAGAGACTATTCTGTGCA 24  
|||||||  
Db 129 CAATTAAAGAGACTATTCTGTGCA 152  
|||||||

RESULT 13

CL680700

LOCUS

DEFINITION

pacificus var. California Pristionchus pacificus genomic, genomic

survey sequence.

ACCESSION CL680700

VERSION CL680700.1

KEYWORDS GSS.

SOURCE Pristionchus pacificus

ORGANISM Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

Neodiplogasteridae; Pristionchus.

1 (bases 1 to 765)

Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.

AppADB: an AcedB database for the nematode satellite organism

Pristionchus pacificus

JOURNAL  
PUBMED  
COMMENT

Nucleic Acids Res. 32 (1), D421-D422 (2004)  
14691447  
Contact: Sommer RJ  
Evolutionary Biology  
Max-Planck-Institute for Developmental Biology  
Spemannstr. 37-39, Tuebingen D-72076, Germany  
Tel: 00497071601371  
Fax: 00497071601498  
Email: ralf.sommer@tuebingen.mpg.de  
This library was generated at Caltech, Pasadena, USA and end  
sequenced at Vancouver, Canada.  
Seq primer: T7  
Class: fosmid ends.

FEATURES  
source

Location/Qualifiers  
1..765  
/organism="Pristionchus pacificus"  
/mol\_type="genomic DNA"  
/strain="California"  
/db\_xref="taxon:54126"  
/clone\_lib="Mixed stage fosmid library of P. pacificus  
var. California"  
/note="Vector: pEpifos-5 Fosmid vector"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 765;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AATTAAGAGACTATTCTGCA 25

Db 698 AATATTGAGATATTCTGCA 721

## RESULT 14

CNS02FKY 785 bp DNA linear GSS 01-SEP-2000  
LOCUS Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone  
DEFINITION 132H08 of library G from Tetraodon nigroviridis, genomic survey  
sequence.

AL195163 1 GI:7833269

GSS; genome survey sequence.

KEYWORDS Tetraodon nigroviridis

SOURCE Tetraodon nigroviridis

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodontidae; Tetraodon.

## REFERENCE

1 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,  
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,  
Saurin,W. and Weissenbach,J.

2 Estimate of human gene number provided by genome-wide analysis  
using Tetraodon nigroviridis DNA sequence

Nat. Genet. 25 (2), 235-238 (2000)

10835645

## REFERENCE

1 Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,  
Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,  
Saurin,W., Bernot,A. and Weissenbach,J.

2 Characterization and repeat analysis of the compact genome of the  
freshwater pufferfish Tetraodon nigroviridis

Genome Res. 10 (7), 939-949 (2000)

10899143

3 (bases 1 to 785)

Genoscope.

Direct Submission

Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

- Web : www.genoscope.cns.fr)

This sequence is a single read and was generated as part of a large  
scale clone-end sequencing project of the Tetraodon nigroviridis  
genome. For more information, please take a look at

## FEATURES

## source

http://www.genoscope.cns.fr/Tetraodon.

Location/Qualifiers

1..785

/organism="Tetraodon nigroviridis"

/mol\_type="genomic DNA"

/db\_xref="taxon:99883"

/clone\_lib="I32H08"

/clone\_lib="G"

/note="Genoscope sequence ID : COAG132DD04SP1

end : PUC-Ori"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 785;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AATTAAGAGACTATTCTGCA 25

Db 111 AATTAAGAGACCAATTGTGCA 134

## RESULT 15

CR146965/c

LOCUS

DEFINITION

CR146965

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..804

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="MHPP345c23"

/clone\_lib="MHPP"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 11; Length 804;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCTGCA 24

Db 188 CAATCAAGAGATTATTCTGCA 165

## RESULT 16

BZ784243/c

LOCUS

DEFINITION

BZ784243

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ784243 810 bp DNA linear GSS 17-MAR-2003  
PUGAT94TB\_ZM\_0\_6\_1\_0\_KB\_Zea\_mays genomic clone ZM9BTa325P20,  
genomic survey sequence.

BZ784243

BZ784243.1 GI:289777840

GSS.

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 810)

**AUTHORS** Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J., and Bennettzen, J.  
**TITLE** Maize Genomics Consortium  
**JOURNAL** Unpublished (2003)  
**COMMENT** Other GSSs: PUGAT94TD  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.

**FEATURES**  
 source  
 1..810  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="ZMMBra325P20"  
 /clone\_lib="ZM\_0.6\_1.0\_KB"  
 /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
 CoT selected genomic DNA library"

**ORIGIN**  
 Query Match 76.8%; Score 19.2; DB 9; Length 810;  
 Best Local Similarity 87.5%; Pred. No. 6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 471 CAAATTAAAGAGACTATTCGTGCA 448

**QY** 1 CAAATTAAAGAGACTATTCGTGCA 24  
 |||||  
 471 CAAATTAAAGAGACTATTCGTGCA 448

**RESULT 17**  
**LOCUS** CR173895/c  
**DEFINITION** Forward strand read from insert in 3'HPRT insertion targeting and chromosome engineering clone MHP183a09, genomic survey sequence.  
**ACCESSION** CR173895  
**VERSION** CR173895.1 GI:49952744  
**KEYWORDS** GSS; genome survey sequence; MICER.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridea; Muridae; Murinae; Mus.  
**REFERENCE** 1 (bases 1 to 812)  
**AUTHORS** Adams, D.J., Biggs, P.J., Cox, A.V., Davies, R.M., van der Weyden, L., Jonkers, J., Smith, J., Plumb, R.W., Taylor, R.G., Nishijima, I., Yu, Y., Rogers, J., and Bradley, A.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (20-FEB-2004) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. http://www.sanger.ac.uk/MICER  
**FEATURES**  
 source  
 1..812  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone\_lib="MHP183a09"  
 /clone\_lib="MHP"

**ORIGIN**  
 Query Match 76.8%; Score 19.2; DB 11; Length 812;  
 Best Local Similarity 87.5%; Pred. No. 6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 809 CAAATCAAGAGATTATTCGAGCA 786

**QY** 1 CAAATTAAAGAGACTATTCGTGCA 24  
 |||||  
 809 CAAATCAAGAGATTATTCGAGCA 786

**RESULT 18**  
**LOCUS** CG348543/c  
**DEFINITION** OGOFU13TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0708D01, genomic survey sequence.  
**ACCESSION** CG348543  
**VERSION** CG348543.1 GI:34265809  
**KEYWORDS** GSS.  
**SOURCE** Zea mays  
**ORGANISM** Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 815)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other GSSs: OGOFU13TH  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TF  
 Class: methylation filtered.  
 Location/Qualifiers  
 1..815  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="ZMMBMA0708D01"  
 /clone\_lib="ZM\_0.7\_1.5\_KB"  
 /note="Vector: pGSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

**ORIGIN**  
 Query Match 76.8%; Score 19.2; DB 10; Length 815;  
 Best Local Similarity 87.5%; Pred. No. 6e+02;  
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 753 CAAATTAAAGAGACTATTCGTGCA 730

**QY** 1 CAAATTAAAGAGACTATTCGTGCA 24  
 |||||  
 753 CAAATTAAAGAGACTATTCGTGCA 730

**RESULT 19**  
**LOCUS** CG348533  
**DEFINITION** OGOFU13TH ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0708D01, genomic survey sequence.  
**ACCESSION** CG348533  
**VERSION** CG348533.1 GI:34265799  
**KEYWORDS** GSS.  
**SOURCE** Zea mays  
**ORGANISM** Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 816)  
 Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other GSSs: OGOFU13TV  
 Contact: Cathy Whitelaw  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843



Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TR  
Class: methylation filtered.  
Location/Qualifiers

## FEATURES

source

1..816  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0708D01"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 10; Length 816;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24

Db 63 CAAATTAAATGGACTATCGGTGCA 86

## RESULT 20

CC681145 872 bp DNA linear GSS 19-JUN-2003  
LOCUS OGUFK42TV ZM 0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0428G11,  
DEFINITION genomic survey sequence.

ACCESSION CC681145

VERSION CC681145.1 GI:32085921

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE Consortium for Maize Genomics

JOURNAL Unpublished (2002)

COMMENT Other GSSs: OGUFK42TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: methylation filtered.

Location/Qualifiers

FEATURES

source

1..872  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0428G11"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 872;  
Best Local Similarity 87.5%; Pred. No. 6e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24

Db 766 CAAATTAAATGGACTATTCGTGCA 789

## RESULT 21

BG479888

LOCUS BG479888

DEFINITION

602527434Fl NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:4650945 5',

mrna sequence.

ACCESSION BG479888

VERSION BG479888.1 GI:13412167

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

REFERENCE 1 (bases 1 to 895)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM1433 row: C column: 10

High quality sequence stop: 771.

Location/Qualifiers

FEATURES

source

1..895

/organism="Homo sapiens"

/mol\_type="mrna"

/db\_xref="taxon:9606"

/clone="IMAGE:4650945"

/tissue\_type="choriocarcinoma"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH MGC 21"

/note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI;

Site 2: EcoRI; cDNA made by oligo-dt priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Query Match 76.8%; Score 19.2; DB 2; Length 895;

Best Local Similarity 87.5%; Pred. No. 6e+02;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24

Db 843 CAAATCAAGTGCCTATTCGTGCA 866

## RESULT 22

CC681138/c

LOCUS CC681138

DEFINITION

OGUFK42TH ZM 0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0428G11,

genomic survey sequence.

ACCESSION CC681138

VERSION CC681138.1 GI:32085914

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 925)

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,

TITLE Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
JOURNAL Consortium for Maize Genomics  
COMMENT Unpublished (2002)  
Other\_GSSs: OGUFK42TV  
Contact: Cathy Whitelaw  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TP  
Class: methylation filtered.

Location/Qualifiers

1..925  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0428G11"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 925;  
Best Local Similarity 87.5%; Pred. No. 6.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTGCA 24  
||||||| ||||| ||||| ||||| |||||

Db 572 CAAATTAAGACTATGCGTGCA 549

## RESULT 23

CC686212/c  
LOCUS OGULK68TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0468K15,  
DEFINITION genomic survey sequence.

ACCESSION CC686212.1 GI:32090988  
VERSION  
KEYWORDS  
SOURCE GSS.

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 991)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
Consortium for Maize Genomics  
Unpublished (2002)

Other\_GSSs: OGUULK68TH

Contact: Cathy Whitelaw  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TP  
Class: methylation filtered.

Location/Qualifiers

1..991  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0468K15"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/note="Vector: pBGSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

Query Match 76.8%; Score 19.2; DB 9; Length 991;  
Best Local Similarity 87.5%; Pred. No. 6.1e+02;  
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTGCA 24  
||||||| ||||| ||||| ||||| |||||

Db 327 CAAATTAAGACTATGCGTGCA 304

## RESULT 24

CG033706/c  
LOCUS PUJFW70TD ZM\_0.6\_1.0\_KB Zea mays genomic clone ZMMBTa0659K19,  
DEFINITION genomic survey sequence.

ACCESSION CG033706  
VERSION CG033706.1 GI:33905862  
KEYWORDS GSS.

SOURCE Zea mays

## ORGANISM

Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 444)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
Bennerzen,J.  
Maize Genomics Consortium  
Unpublished (2003)  
Other\_GSSs: PUJFW70TB  
Contact: Cathy Whitelaw  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TP  
Class: sheared ends.

## FEATURES

source  
1..444  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBTa0659K19"  
/clone\_lib="ZM\_0.6\_1.0\_KB"  
/note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
CoT selected genomic DNA library"

## ORIGIN

Query Match 75.2%; Score 18.8; DB 10; Length 444;  
Best Local Similarity 90.9%; Pred. No. 8.9e+02;  
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTCTGTGCA 24  
||||||| ||||| ||||| ||||| |||||

Db 437 AATTAAGAGACTATTGTGAA 416

## RESULT 25

AO989224/c  
LOCUS Gm\_ISB001\_053\_H20F ISU Soybean BAC Library (pBelBAC11 HindIII)  
DEFINITION Glycine max genomic clone Gm\_ISB001\_053\_H20; genomic survey  
sequence.

ACCESSION AO989224  
VERSION AO989224.1 GI:6937170  
KEYWORDS GSS.

## ORGANISM

Glycine max (soybean)  
Glycine max  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
Glycine.

REFERENCE  
 AUTHORS Marek,L.F., Paz,M., Darnielle,L., Hanson,N. and Shoemaker,R.C.  
 TITLE BAC End sequences from a soybean genomic library (ISU)  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Shoemaker Randy C  
 Agronomy Department  
 Iowa State University  
 Ames, IA 50011-1010, USA  
 Tel: 515 294 1205  
 Fax: 515 294 2299  
 Email: rcsshoe@iastate.edu  
 This BAC identified by SSR Satt596. For more information, see  
 Soybase at:  
<http://genome.cornell.edu/cgi-bin/WebBase/webase?db=soybase>.  
 Please see as an authority for the mapping/naming: Cregan P.B., T.  
 Jarvik, A.L., Bush, R.C. Shoemaker, K.G. Lark, A.L. Kahler, N. Kaya,  
 T.T. VanRoai, D.G. Lohnes, J. Chung, and J.E. Specht. 1999a. An  
 integrated genetic linkage map of the soybean genome. Crop Sci.  
 39: (in press)  
 Seq primer: M13F  
 Class: BAC ends

FEATURES  
 source  
 1. .460  
 /organism="Glycine max"  
 /mol\_type="genomic DNA"  
 /cultiivar="Williams82"  
 /db\_xref="taxon:3847"  
 /clone="Gm15B001\_053\_H20;"  
 /tissue\_type="primary leaves"  
 /dev\_stage="cotyledon stage"  
 /clone\_lib="ISU Soybean BAC Library (pBelobAC11 HindIII)"  
 /notes="Vector: pBelobAC11; The ISU BAC library (Marek and Shoemaker, Genome 40:420, 1997) was constructed using the HindIII site in pBelobAC11. The library consists of approximately 40,000 clones with an average insert size of 150 Kb, equal to 4 haploid genome equivalents. Screening of the library is done by PCR amplification of DNA pools."

ORIGIN  
 Query Match 75.2%; Score 18.8; DB 9; Length 460;  
 Best Local Similarity 90.9%; Pred. No. 8.e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AATTAAGAGACTATTCGTGC 23  
 |||||  
 Db 297 AATTAAGAGACAATTAGTGC 276

RESULT 26  
 AQ233842  
 LOCUS 612 bp DNA linear GSS 29-SEP-1998  
 DEFINITION HS 2057 A1 B09 T7 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2057 Col=17 Row=C, genomic survey sequence.

ACCESSION AQ233842  
 VERSION AQ233842.1 GI:3662449  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 612)  
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 PUBMED 10449764  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High throughput Sequencing Center

University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Sequence Tagged Connector  
 Plate: 2057 row: C column: 17  
 Class: BAC ends  
 High quality sequence stop: 612.

#### FEATURES

source

1. .612  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=2057 Col=17 Row=C"  
 /sex="male"  
 /clone\_lib="CIT Approved Human Genomic Sperm Library D"  
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

#### ORIGIN

Query Match 75.2%; Score 18.8; DB 9; Length 612;  
 Best Local Similarity 90.9%; Pred. No. 9.e+02;  
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 3 AATTAAGAGACTATTCGTGCA 24  
 |||||  
 Db 424 AATTAAGATACATTCATGCA 445

#### RESULT 27

CV709364

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Poncirus trifoliata

Poncirus trifoliata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

REFERENCE 1 (bases 1 to 148)

AUTHORS

Roose,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,

Wanmaker,S., Kim,H.R., Kudrna,D. and Stum,D., Wlssotski,M.,

Wing,R.

Development of EST Resources and New Genetic Markers for California

Citrus - Poncirus trifoliata CTV-challenged phloem - AGI2

Unpublished (2004)

CONTACT: Mikeal Roose

Department of Botany & Plant Sciences, University of California

Riverside, CA, 92521-0124, USA

Tel: 9097874137

Fax: 9097874437

Email: mikeal.roose@ucr.edu

Seq primer: T7.

Location/Qualifiers

1. .148

/organism="Poncirus trifoliata"

/mol\_type="mRNA"

/cultiivar="Pomeroy Op"

/db\_xref="taxon:37690"

/clone="PT\_6Ea0012M03"

/tissue\_type="Phloem"

/dev\_stage="10 - 30 cm shoots"

/lab\_host="E. coli TJCl21"

/clone\_lib="Poncirus trifoliata CTV-challenged cdna

library - AGI2"

/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;

Site 1: EcoRI; Site 2: XhoI; Plants were grown in the

greenhouse at University of California, Riverside. The

scion was a open-pollinated (very probably selfed)

seedling of *Poncirus trifoliata* cv Pomeroy that was selected as homozygous for the Ctv resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate #514 over 1 year before sampling (CTV infects sweet orange, but not genotypes carrying the Ctv resistance gene). Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA library was made, and 0.5 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript SK(-) phagemids. All steps to this point were performed in the MLRoose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files were downloaded to UC Riverside (Close), then processed at UC Riverside (Wanamaker) using the HarVESt pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 148;  
Best Local Similarity 84.0%; Pred. No. 1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCGTGCAA 25

Db 115 CAATTAATACAATATCGTGCAA 139

## RESULT 28

AZ794021 206 bp DNA linear GSS 16-FEB-2001  
LOCUS  
DEFINITION  
2M0047B07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic  
clone UUGC2M0047B07 R, genomic survey sequence.

ACCESSION AZ794021

VERSION 1

KEYWORDS GSS.

SOURCE AZ794021.1 GI:12939565

ORGANISM Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
Niederhausern,A. and Wright,D.,Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts

## TITLE

## JOURNAL

## COMMENT

Unpublished (2000)  
Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0047 row: B column: 07

Seq primer: CACACGAGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 206.

Location/Qualifiers

1..206

/organism="Mus musculus"

/mol\_type="genomic DNA"

## FEATURES

## source

/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0047B07"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC1M library"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(<http://www.jax.org/resources/documents/dnares/>). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adaptored DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gi|4732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adaptored mouse DNA was annealed to  
adaptored vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 206;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCGTGCAA 25

Db 132 CAATTAATAACACTATTTTTCGAA 156

## RESULT 29

DE043516 537 bp DNA linear GSS 25-MAY-2005  
LOCUS  
DEFINITION  
Oryzias latipes DNA, clone: olal-006G19.F, genomic survey sequence.

ACCESSION DE043516

VERSION DE043516.1

KEYWORDS GSS.

SOURCE DE043516.1 GI:62555060

ORGANISM Oryzias latipes (Japanese medaka)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1  
Fujiyama, A., Toyoda, A., Kuroki, Y. and Sakaki, Y.  
BAC end sequences of Olal Oryzias latipes Library  
Published Only in Database (2005)  
2 (bases 1 to 537)  
Fujiyama, A.  
Direct Submission  
Submitted (12-APR-2005) Asao Fujiyama, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);  
1-7-22 Suehiro-chou, Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
(E-mail: [afujiyam@gsc.riken.jp](mailto:afujiyam@gsc.riken.jp) URL: <http://att.gsc.riken.jp/>,  
Tel:81-3-4212-2558, Fax:81-3-3556-1916)  
This work was done in collaboration with Takeda, H. (1), Naruse, K.  
(2)

and Narita, T. (3)

(1) Department of Biological Science,

University of Tokyo

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN

Phone: +81-3-5841-4431

Fax: +81-3-5841-4993

E-mail: [htakeda.s.u-tokyo.ac.jp](mailto:htakeda.s.u-tokyo.ac.jp)

(2) Department of Biological Science,

University of Tokyo

Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN

Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: naruse.s.u-tokyo.ac.jp  
 (3) Department of Biological Science,  
 University of Tokyo  
 Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, JAPAN  
 Phone: +81-3-5841-4431  
 Fax: +81-3-5841-4993  
 E-mail: tanarita.s.u-tokyo.ac.jp

## PRIMERS

Sequencing : Forward

## LIBRARY

Vector : pK3145  
 R.Site 1 : SacI  
 L.Site 2 : SacI  
 Location/Qualifiers  
 1. .537  
 /organism="Oryzias latipes"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:8090"  
 /clone="Olai-006G19.F"  
 /sex="male"  
 /cell\_type="whole body"  
 /clone\_lib="BAC end sequences of Olai Oryzias latipes  
 library"

FEATURES  
source

## ORIGIN

Query Match 74.4%; Score 18.6; DB 11; Length 537;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
 |||||  
 DB 203 CAAATTACAGAGAAATACGTGCAA 227

## RESULT 30

BZ615728 569 bp DNA linear GSS 16-JAN-2003  
 1954e11.g1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone  
 1954e11 5', genomic survey sequence.

ACCESSION BZ615728  
 VERSION BZ615728  
 KEYWORDS GSS.  
 SOURCE Zea mays

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

AUTHORS Rabinowicz,P.D., O'Shaughnessy,A.L., Ballija,V., Dedhia,N.,  
 Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,  
 Zutavern,T., McCombie,W.R. and Martienssen,R.A.  
 Genomic shotgun sequences from Zea mays (methyl-filtered)

## TITLE

JOURNAL Unpublished (2002)  
 COMMENT Contact: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874

Email: mcombie@cshl.org  
 Plate: 1954 row: e column: 11  
 Seq primer: -21M13UnivRev  
 Class: shotgun

High quality sequence stop: 569.

FEATURES  
source

Location/Qualifiers  
 1. .569  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /cultiivar="B73"  
 /db\_xref="taxon:4577"  
 /clone="1954e11"

/lab\_host="DH5a"  
 /clone\_lib="WGS-ZmaysF (DH5a methyl filtered)"  
 /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;  
 The vector was digested with XbaI and one nucleotide was  
 added by fill in the recessive 3' end. The genomic DNA  
 was nebulized, end repaired, adaptor ligated and size  
 fractionated using sephadex. The resulting fragments were  
 between 0.8 and 3 kb and were cloned into the vector (X/Y  
 reads in M13mpl9, b/g reads in pUC19). The same ligation  
 was transformed into DH5a."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 569;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
 |||||  
 DB 100 CAAATGAACAGACTACTCTGTCAA 124

## RESULT 31

CB092313 586 bp mRNA linear EST 27-JAN-2003  
 hf03h07.g1 Cycad Leaf Library (NYBG) Cycas rumphii cDNA clone  
 hf03h07, mRNA sequence.

ACCESSION CB092313.1 GI:27916505  
 VERSION CB092313  
 KEYWORDS EST.  
 SOURCE Cycas rumphii

## ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Cycadophyta; Cycadales; Cycadaceae; Cycas.

## REFERENCE

AUTHORS Brenner,E.D., Katari,M.S., Dedhia,N.N., O'Shaughnessy,A.L.,  
 Ballija,V., Martienssen,R.A., McCombie,R.W., Benfey,P., Coruzzi,G.  
 and Stevenson,D.  
 Expressed tag sequences from Cycas ovule (NYBG)

## TITLE

JOURNAL Unpublished (2003)

COMMENT Contact: W. Richard McCombie  
 Lita Annenberg Hazen Genome Sequencing Center  
 Cold Spring Harbor Laboratory  
 PO Box 100, Cold Spring Harbor, NY 11724, USA  
 Tel: 516 367 8884  
 Fax: 516 367 8874  
 Email: mcombie@cshl.org  
 Plate: hf03 row: h column: 07  
 Seq primer: -21M13UnivRev  
 High quality sequence stop: 586.

## FEATURES

Location/Qualifiers  
 1. .586  
 /organism="Cycas rumphii"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:58031"  
 /clone="hf03h07"  
 /sex="Female"

/clone\_lib="Cycad Leaf Library (NYBG)"

/notes="Organ: Young leaf; Vector: pBK-CMV; Site 1: Xho I;  
 Site 2: Eco RI; Date: Completed 09/01/2001. Submitted to  
 CSHL 09/05/2001. Sample: Young emergent leaves. From New  
 York Botanical Garden Conservatory accession number 808/59  
 A (collected 03/2001). Library: Made using Stratagene's  
 ZAP Express Vector Kit. Library was size fractionated for  
 large inserts."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 6; Length 586;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCAA 25  
 |||||  
 DB 485 CAAATGAGAGAGGCTATTCTGTCAA 509

## ORIGIN

```

RESULT 32
CX543634
LOCUS
DEFINITION
  UCRPT01_5_008 B05 T7 Poncirus trifoliata CTV-challenged cDNA
  library - UCRPT01_UCR2 Poncirus trifoliata cDNA clone
  UCRPT01_008_T7_B05, mRNA sequence.
ACCESSION
CX543634
VERSION
CX543634.1 GI:57570659
KEYWORDS
EST.
SOURCE
Poncirus trifoliata
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
REFERENCE
1 (bases 1 to 600)
  Roose,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,
  Wanamaker,S., Choi,Y. and Kingan,T.
  Development of EST Resources and New Genetic Markers for California
  Citrus - Poncirus trifoliata CTV-Challenged Phloem - UCRPT01-UCR2
  Unpublished (2004)
  Contact: Mikeal Roose
  Department of Botany & Plant Sciences, University of California
  Riverside, CA, 92521-0124, USA
  Tel: 9097874137
  Fax: 9097874437
  Email: mikeal.roose@ucr.edu
  Seq primer: T7.
FEATURES
    Location/Qualifiers
      1..600
        /organism="Poncirus trifoliata"
        /mol_type="mRNA"
        /cultivar="Pomeroy OP"
        /db_xref="taxon:37690"
        /clone="UCRPT01_008_T7_B05"
        /tissue_type="bark (with phloem)"
        /dev_stage="10 - 30 cm shoots"
        /lab_host="E. coli TUC121"
        /clone_lib="Poncirus trifoliata CTV-challenged cDNA
        library - UCRPT01-UCR2"
        /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
        Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
        greenhouse at University of California, Riverside. The
        scion was an open-pollinated (very probably selfed)
        seedling of Poncirus trifoliata cv Pomeroy that was
        selected as homozygous for the Ctv resistance gene. The
        rootstock was sweet orange infected with citrus tristeza
        virus (CTV) isolate T514 over 1 year before sampling (CTV
        infects sweet orange, but not genotypes carrying the Ctv
        resistance gene. Shoots 10-30 cm long were harvested in
        October 2000, and the green phloem (bark) was removed and
        frozen quickly in dry ice. Total RNA was extracted using
        TriZol reagent (Gibco). Poly(A) RNA was purified, a cDNA
        library was made, and 0.5 million primary lambda cDNA
        clones were in vivo excised to give a population of
        Bluescript SK(-) phagemids. All steps to this point were
        performed in the ML Roose lab at the University of
        California, Riverside by X. Ye. Phagemids were plated,
        plasmid DNA purified, cDNA clones archived, and DNA
        sequences determined bi-directionally using an ABI3730 at
        the University of California Riverside Institute of
        Integrative Genome Biology Genomics Core Instrumentation
        Facility, (Choi, Kingan). Chromatogram files were
        downloaded by FTP by Close, then processed by Wanamaker
        (Close lab) using the HarVest pipeline
        (http://harvest.ucr.edu) to remove vector and cloning
        oligo sequences and various contaminants, and to trim to a
        high quality region. Sequences that retained a phred 17
        region of at least 100 bases were assembled, then chimeras
        were removed following manual inspection of assemblies
        (Close, Roose, Wanamaker). Sequences that survived all
        removal steps were submitted to GenBank."

```

## ORIGIN

```

Query Match      74.4%; Score 18.6; DB 8; Length 600;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      1 CAAATTAAGAGACTATTTCGTGCA 25
      ||| ||| ||| ||| ||| ||| |||
Db     114 CAAATTAATACATATCCGTGCAA 138

RESULT 33
CX644973/c
LOCUS
DEFINITION
  UCRPT02_8F01_g Poncirus trifoliata Roots with Iron Deficiency -
  UCRPT02-UCR1 Poncirus trifoliata cDNA clone UCRPT02-8F01-L1-1-5.9,
  mRNA sequence.
ACCESSION
CX644973
VERSION
CX644973.1 GI:57879802
KEYWORDS
EST.
SOURCE
Poncirus trifoliata
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
REFERENCE
1 (bases 1 to 604)
  Close,T.J., Roose,M.L., Parker,D.R., Federici,C.F., Mandal,J.,
  Fenton,R.D., Wanamaker,S., Landry,B., Hubert,N., Laforest,M.,
  Landry,J. and Ligonde,A.
  Development of EST Resources and New Genetic Markers for California
  Citrus - Poncirus trifoliata Roots with Iron Deficiency -
  UCRPT02-UCR1
  Unpublished (2005)
  Contact: Timothy J. Close
  Department of Botany & Plant Sciences
  University of California
  Riverside, CA 92521-0124, USA
  Tel: 909-787-3318
  Fax: 909-787-4437
  Email: timothy.close@ucr.edu
  Seq primer: T3.
FEATURES
    Location/Qualifiers
      1..604
        /organism="Poncirus trifoliata"
        /mol_type="mRNA"
        /cultivar="Pomeroy"
        /db_xref="taxon:37690"
        /clone="UCRPT02-8F01-L1-1-5.9"
        /tissue_type="root"
        /dev_stage="seedling"
        /lab_host="E. coli TUC121"
        /clone_lib="Poncirus trifoliata Roots with Iron Deficiency
        - UCRPT02-UCR1"
        /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
        Site 1: EcoRI; Site 2: XhoI; Plant materials were prepared
        138 days after sowing were about 20-23 cm in height,
        bearing 8-12 leaves. On May 26, 2004, plants were washed
        free of soil with a stream of water, then placed upright
        with the roots submerged in two tanks of nutrient
        solution. The solution was sufficient in all major and
        minor nutrients and buffered with MES at about pH 6.95.
        The plants were maintained in this until June 14, 2004,
        when the solution was changed to one with only 20
        micromolar iron, chelated with EDTA to induce iron
        deficiency. The pH was maintained at 6.99 by sodium
        carbonate/CO2 buffering. This solution was replaced on
        July 6, 2004. Roots from three plants were sampled on
        June 16, June 21, July 1 and July 26, 2004. Roots were
        collected by removing the plant from the nutrient
        solution, blotting off excess moisture with a paper towel,
        then cutting off the top of the plants. Three plants were
        pooled in one aluminum foil packet, and frozen between two

```

sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using Trizol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wanamaker (Close lab) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a trim 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roese, Wanamaker). Sequences that survived all removal steps were submitted to GenBank."

ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 604;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGACATATTCGTGCAA 25

Db 417 CAAATTAAATCAATATTCGTGCAA 393

RESULT 34  
CX644972 608 bp mRNA linear EST 18-JAN-2005  
LOCUS UCRPT02-8F01\_b Poncirus trifoliata Roots with Iron Deficiency -  
DEFINITION UCRPT02-UCR1\_Poncirus trifoliata cDNA clone UCRPT02-8F01-L1-1-6.b,  
mRNA sequence.

ACCESSION CX644972.1 GI:57879801

VERSION CX644972

KEYWORDS EST.

SOURCE Poncirus trifoliata

ORGANISM Poncirus trifoliata

REFERENCE Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AUTHORS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

TITLE rosid; eurosids II; Sapindales; Rutaceae; Poncirus.

1 (bases 1 to 608)

Close,T.J., Roese,M.L., Parker,D.R., Federici,C.F., Mandal,J.,

Fenton,R.D., Wanamaker,S., Landry,B., Hubert,N., Laforest,M.,

Landry,J. and Ligonde,A.

Development of EST Resources and New Genetic Markers for California

Citrus - Poncirus trifoliata Roots with Iron Deficiency -

UCRPT02-UCR1

FEATURES  
Source

Email: timothy.close@ucr.edu  
Seq primer: 17.  
Location/Qualifiers  
1..608  
/organism="Poncirus trifoliata"  
/mol\_type="mRNA"  
/cultivar="Pomeroy"  
/db\_xref="taxon:37690"  
/clone="UCRPT02-8F01-L1-1-6.b"  
/tissue\_type="root"  
/dev\_stage="seedling"  
/lab\_host="E. coli TUC121"  
/clone\_lib="Poncirus trifoliata Roots with Iron Deficiency - UCRPT02-UCR1"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; plant materials were prepared by Federici (Roese lab) with advice from Parker. Seedlings 138 days after sowing were about 20-23 cm in height, bearing 8-12 leaves. On May 26, 2004, plants were washed free of soil with a stream of water, then placed upright with the roots submerged in two tanks of nutrient solution. The solution was sufficient in all major and minor nutrients and buffered with MES at about pH 6.95. The plants were maintained in this until June 14, 2004, when the solution was changed to one with only 20 micromolar iron, chelated with EDTA to induce iron deficiency. The pH was maintained at 6.99 by sodium carbonate/CO2 buffering. This solution was replaced on July 6, 2004. Roots from three plants were replaced on June 16, June 21, July 1 and July 28, 2004. Roots were collected by removing the plant from the nutrient solution, blotting off excess moisture with a paper towel, then cutting off the top of the plants. Three plants were pooled in one aluminum foil packet, and frozen between two sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80°C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of *Thielaviopsis brevicola*, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using Trizol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wanamaker (Close lab) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a trim 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roese, Wanamaker). Sequences that survived all removal steps were submitted to GenBank."

ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 608;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;







	Matches	21;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
1	CAAAATTAAGAGACTATTCGTGCAA	25								
536	CATATTAAAGGAGCATATTCGTGCAA	512								

RESULT 38  
W333363/c  
AW333363 691 bp mRNA linear EST 31-JAN-2000  
S20G9 AGS-1 Pneumocystis carinii CDNA 3', mRNA sequence.

VERSION AW333363.1 GI:6829720  
 SYWORDS EST.  
 SOURCE Pneumocystis carinii  
 ORGANISM Pneumocystis carinii  
 Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes;  
 Pneumocystidaceae; Pneumocystis.

1. Issues 1 to 2011

**REFERENCE**  
**AUTHORS**  
Smulian, A.G., Arnold, J., Weise, M., Wunderlich, J., Staben, C.,  
Edman, J.C., Kovacs, J. and Cushion, M.  
**TITLE**  
Expressed sequence tags from *Pneumocystis carinii*  
**JOURNAL**  
Unpublished (2000)  
**COMMENT**  
Contact: Staben C  
School of Biological Sciences  
University of Kentucky  
101 Morgan Building, University of Kentucky, Lexington, KY  
40506-0225, USA  
Tel: 606 257 2161  
Fax: 606 257 1717  
Email: [staben@pop.uky.edu](mailto:staben@pop.uky.edu).

FEATURES	source
Location/Qualifiers	
1..691	
/organism="Pneumocystis carinii"	
/mol_type="mRNA"	
/db_xref="taxon:4754"	
/lab_host="E. coli"	
/clone_lib="AGS-1"	
/note="Vector: Lambda ZAP II; Site 1: EcoRI; Site 2: XhoI; P. carinii organisms (3x10e9) from a single rat [99-1-6, sacrificed on 3/17/99] at Cincinnati VA facilities. Trizol extracted RNA. Oligo dt priming, standard conditions described by vendor, Stratagene. Further details see <a href="http://www.ukv.edu/project/Pneumocystis/">www.ukv.edu/project/Pneumocystis/</a> "	

Query Match	74.4%	Score 18.6;	DB 1;	Length 691;
Best Local Similarity	84.0%;	Pred. NO. 1.1e+03;		
Matches 21;	Conservative	0;	Mismatches 4;	Indels 0;
Gaps 0;				

1 CAAATTAAAGAGACTATTTCGTGCAA 25  
|||||  
619 CAAATTAAAGAGAAAAATTCGTGCAA 595

RESULT 39		706 bp	linear	EST 12-JAN-2005
XX546303	UCRPT01_5_028_B02_T7	Poncirus trifoliata	CTV-challenged	CDNA
XX546303	library - UCRPT01-UCR2	Poncirus trifoliata	CDNA clone	
XX546303	UCRPT01_028_T7_B02_	mRNA sequence.		

CESSION  
CX546303  
CX546303.1  
GI:57573328  
EST.  
EYWORDS

SOURCE	ORGANISM
Poncirus trifoliata	Poncirus trifoliata
Poncirus trifoliata	Poncirus trifoliata
Eukarya; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; rosids; eurosids II; Sapindales; Rosaceae; Poncirus.	

REFERENCE	1 (bases 1 to 706)
AUTHORS	Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D., Wanamaker, S., Choi, Y. and Kingan, T.
TITLE	Development of EST resources and New Genetic Markers for California

JOURNAL  
COMMENT

Cirrus - *Poncirus trifoliata* CTV-challenged phloem - UCRPT01-UCR2  
Unpublished (2004)  
Contact: Mikeal Roose  
Department of Botany & Plant Sciences, University of California  
Riverside, CA, 92521-0124, USA  
Tel: 9097874137  
Fax: 9097874437  
Email: mikeal.roose@ucr.edu  
Seq primer: T7.

## FEATURES

```

1. .706
/organism="Poncirus trifoliata"
/mol_type="mRNA"
/cultivar="Pomeroy OP"
/db_xref="taxon:37690"
/clone="UCRP01.028.T7.B02"
/tissue_type="bark [with phloem]"
/dev_stages="10 - 30 cm shoots"
/lab_host="E. coli TUC121"
/clone_lib="Poncirus trifoliata CTV-challenged cDNA
library - UCRP01-UCR2"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
Greenhouse at University of California, Riverside. The
scion was an open-pollinated (very probably selfed)
seedling of Poncirus trifoliata cv Pomeroy that was
selected as homozygous for the CTV resistance gene. The
rootstock was sweet orange infected with citrus tristeza
virus (CTV) isolate TS14 over 1 year before sampling (CTV
infects sweet orange, but not genotypes carrying the CTV
resistance gene. Shoots 10-30 cm long were harvested in
October 2000, and the green phloem (bark) was removed and
frozen quickly in dry ice. Total RNA was extracted using
Trizol reagent (Gibco). Poly(A) RNA was purified, a cDNA
library was made, and 0.5 million primary lambda cDNA
clones were in vivo excised to give a population of
phagescript SK(-) phagemids. All steps to this point were
performed in the MLRoose lab at the University of
California, Riverside by X. Ye. Phagemids were plated,
plasmid DNA purified, cDNA clones archived, and DNA
sequences determined bi-directionally using an ABI3730 at
the University of California Riverside Institute of
Integrative Genome Biology Genomics Core Instrumentation
Facility. (Choi, Kingan). Chromatogram files were
downloaded by FTP by Close, then processed by Wanamaker
(Close lab) using the HarVEST pipeline
(http://harvest.ucr.edu) to remove vector and cloning
oligo sequences and various contaminants, and to trim to a
high quality region. Sequences that retained a phred 17
region of at least 100 bases were assembled, then chimeras
were removed following manual inspection of assemblies
(Close, Roose, Wanamaker). Sequences that survived all
removal steps were submitted to GenBank."
```

## ORIGIN

Query Match	74.4%;	Score 18.6;	DB 8;	Length 706;
Best Local Similarity	84.0%;	Pred. No. 1.1e+03;		
Matches	21.	Conservative	0;	Mismatches 4;
			Indels	0;
			Gaps	0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
|||||  
90 CAAATTAAATACAATATCCGTGCAA 114  
|||||

RESULT 40  
CJ397741/C

CJ397741 708 bp mRNA linear EST 21-JUN-2005  
CJ397741 *Molgula tectiformis* unpublished cDNA library, gonad  
*Molgula tectiformis* cDNA clone mtgda001c07 3', mRNA sequence.  
CJ397741  
CJ397741.1 GI:68089611  
EST.  
*Molgula tectiformis*

```

ORGANISM      Molgula tectiformis
               Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
               Stolidobranchia; Molgulidae; Molgula.
REFERENCE      1 (bases 1 to 708)
AUTHORS        Gyoja, F., Satou, Y. and Satoh, N.
TITLE          Expressed genes in Molgula tectiformis
JOURNAL        Unpublished (2005)
COMMENT        Contact: Tadashi Shin-i
               Center For Genetic Resource Information
               National Institute of Genetics
               1111 Yata, Mishima, Shizuoka 411-8540, Japan
               Tel: 81-559-81-6856
               Fax: 81-559-81-6855
               Email: tshin@genes.nig.ac.jp
               When you want to obtain this EST clone, please send an e-mail to
               Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka
               Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).
FEATURES       Location/Qualifiers
               source
               1..708
               /organism="Molgula tectiformis"
               /mol_type="mRNA"
               /db_xref="taxon:30286"
               /clone="mtgd001c07"
               /tissue_type="gonad"
               /clone_lib="Molgula tectiformis unpublished cDNA library,
               gonad"
ORIGIN
Query Match      74.4%; Score 18.6; DB 7; Length 708;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  CAAATTAAAGAGACTATTCTGCAAA 25
        ||||| ||||| ||||| ||||| |||||
DB      57  CAAATTATAGAGATTATTCTGCAAA 33

RESULT 41
LOCUS      CV509866/c 712 bp mRNA linear EST 05-OCT-2004
DEFINITION kc52a06.y1 Xiphinema index CSEQDA01 Xiphinema index cDNA 5' similar
            to TR-0622295 O62295 Jc8.6B PROTEIN. [2] WP:CE17989 ;, mRNA
            sequence.
ACCESSION  CV509866
VERSION     CV509866.1 GI:53807979
KEYWORDS    EST.
SOURCE      Xiphinema index
            Xiphinema index
            Eukaryota; Metazoa; Nematoda; Enoplea; Dorylaimida; Dorylaimina;
            Longidoroidea; Longidoridae; Xiphinema.
REFERENCE   1 (bases 1 to 712)
            McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J.,
            Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,
            Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C.,
            Tsagariswili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,
            Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T.,
            Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,
            McCann, R., Waterston, R. and Wilson, R.
            The Washington Univ. Nematode EST Project, 1999
            Unpublished (1999)
            Contact: Mitreva M
            The Washington Univ. Nematode EST Project, 2004
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@wustl.wustl.edu
            Library materials provided by: J.T. Jones Library constructed by:
            Incyte Genomics (for J.T. Jones)
            Seq primer: T7 primer
            High quality sequence stop: 691.
FEATURES       Location/Qualifiers
               source
               1..712

```

```

/organism="Xiphinema index"
/mol_type="mRNA"
/db_xref="taxon:46003"
/sex="mixed"
/dev_stage="mixed"
/lab_host="GC10 (only ligation mix supplied)"
/clone_lib="Xiphinema index CSEQDA01"
/note="Vector: pSport1; Site 1: 5'; Sali; Site 2: 3';
NOTE: The library was made from mRNA isolated from total
RNA using oligo dt cellulose. Total RNA was generated from
mixed stage X. index extracted from soil around fig (Ficus
serica) roots. Nematodes were cleaned by passing through a
series of sieves. Pure X. index were obtained by hand
sorting of nematodes. The library was constructed by
Incyte Genomics by cloning cDNA directionally into the
pSport1 vector using Sali and NotI sites. The library was
provided by John T. Jones (jones@scri.sari.ac.uk) at
SCRI, Nematology Department. Sequencing by: Washington
University Genome Sequencing Center, St. Louis, MO."
ORIGIN
Query Match      74.4%; Score 18.6; DB 7; Length 712;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  CAAATTAAAGAGACTATTCTGCAAA 25
        ||||| ||||| ||||| ||||| |||||
DB      505  CAAATTATTGTGACAAATTCGTGCAA 481

RESULT 42
LOCUS      CZ287609 712 bp DNA linear GSS 01-JUL-2005
DEFINITION cp53ell.r Candida parapsilosis Random Genomic Library Candida
            parapsilosis genomic clone cp53ell, genomic survey sequence.
ACCESSION  CZ287609
VERSION     CZ287609.1 GI:68458919
KEYWORDS    GSS.
SOURCE      Candida parapsilosis
            Candida parapsilosis
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1 (bases 1 to 712)
            Logue, M.E., Wong, S., Wolfe, K.H. and Butler, G.
            A genome sequence survey shows that the pathogenic yeast Candida
            parapsilosis has a defective MTL1 allele at its mating type locus
            Eukaryot. Cell 4 (6), 1009-1017 (2005)
            15947193
            Contact: Logue M
            Department of Biochemistry, Conway Institute of Biomolecular and
            Biomedical Research
            University College Dublin
            Dublin 4, Ireland
            Tel: +353 1 7166885
            Fax: +353 1 2837211
            Email: mary.e.logue@ucd.ie
            Class: plasmid ends.
            Location/Qualifiers
               source
               1..712
               /organism="Candida parapsilosis"
               /mol_type="genomic DNA"
               /strain="CLIB214"
               /db_xref="taxon:5480"
               /clone="cp53ell"
               /clone_lib="Candida parapsilosis Random Genomic Library"
ORIGIN
Query Match      74.4%; Score 18.6; DB 10; Length 712;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  CAAATTAAAGAGACTATTCTGCAAA 25
        ||||| ||||| ||||| ||||| |||||

```

### Query Match

high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 727;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAGAGACTATTCGTGCAA 25  
|||||  
Db 156 CAAATTAACATATATCGTGCAA 180  
|||||

## RESULT 45

CJ344248/c 731 bp mRNA linear EST 14-JUN-2005  
LOCUS CJ344248 Molgula tectiformis unpublished cDNA library Molgula  
tectiformis cDNA clone mtbh025m12 3', mRNA sequence.  
ACCESSION CJ344248

VERSION CJ344248.1 GI:67751297  
KEYWORDS EST.  
SOURCE Molgula tectiformis

ORGANISM Molgula tectiformis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
Stolidobranchia; Molgulidae; Molgula.

## REFERENCE

AUTHORS Gyoja, F., Satou, Y. and Satoh, N.

TITLE Expressed genes in Molgula tectiformis

JOURNAL Unpublished (2005)

COMMENT Contact: Tadao Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp

When you want to obtain this EST clone, please send an e-mail to  
Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka  
Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).

## FEATURES

source

1..731  
/organism="Molgula tectiformis"  
/mol\_type="mRNA"  
/db\_xref="taxon:30286"  
/clone="mtbh025m12"  
/tissue\_type="whole animal"  
/dev\_stage="embryo just before hatching"  
/clone\_lib="Molgula tectiformis unpublished cDNA library"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 731;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAGAGACTATTCGTGCAA 25  
|||||  
Db 44 CAAATTATAGAGATTATTCGCAA 20  
|||||

## RESULT 46

CJ357933/c 732 bp mRNA linear EST 15-JUN-2005  
LOCUS CJ357933 Molgula tectiformis unpublished cDNA library, cleaving  
embryo Molgula tectiformis cDNA clone mtel006f20 3', mRNA sequence.  
ACCESSION CJ357933

VERSION CJ357933.1 GI:67798081  
KEYWORDS EST.  
SOURCE Molgula tectiformis

ORGANISM Molgula tectiformis  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
Stolidobranchia; Molgulidae; Molgula.

## REFERENCE

AUTHORS Gyoja, F., Satou, Y. and Satoh, N.

TITLE  
JOURNAL  
COMMENT

Expressed genes in Molgula tectiformis  
Unpublished (2005)  
Contact: Tadao Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshin@genes.nig.ac.jp  
When you want to obtain this EST clone, please send an e-mail to  
Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka  
Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).

## FEATURES

source

1..732  
/organism="Molgula tectiformis"  
/mol\_type="mRNA"  
/db\_xref="taxon:30286"  
/clone="mtel006f20"  
/tissue\_type="whole animal"  
/dev\_stage="cleaving embryo"  
/clone\_lib="Molgula tectiformis unpublished cDNA library,  
cleaving embryo"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 732;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 CAAATTAAGAGACTATTCGTGCAA 25  
|||||  
Db 59 CAAATTATAGAGATTATTCGCAA 35  
|||||

## RESULT 47

CX641978 733 bp mRNA linear EST 18-JAN-2005  
LOCUS CX641978 UCRPT02-UCR1 Poncirus trifoliata Roots with Iron Deficiency -  
DEFINITION UCRPT02-UCR1 Poncirus trifoliata cDNA clone UCRPT02-57B01-D2-1-6.b,  
mRNA sequence.  
ACCESSION CX641978  
VERSION CX641978.1 GI:57876807  
KEYWORDS EST.  
SOURCE Poncirus trifoliata

## ORGANISM

Poncirus trifoliata  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

## REFERENCE

AUTHORS

Close, T.J., Roose, M.L., Parker, D.R., Federici, C.F., Mandal, J.,  
Penton, R.D., Wanamaker, S., Landry, B., Hubert, N., Laforest, M.,  
Landry, J. and Ligonde, A.

Development of EST Resources and New Genetic Markers for California  
Citrus - Poncirus trifoliata Roots with Iron Deficiency -  
UCRPT02-UCR1

## JOURNAL

COMMENT

Contact: Timothy J. Close  
Department of Botany & Plant Sciences  
University of California  
Riverside, CA 92521-0124, USA  
Tel: 909-787-3318  
Fax: 909-787-4437  
Email: timothy.close@ucr.edu

Seq primer: T7.

## FEATURES

source

1..733  
/organism="Poncirus trifoliata"  
/mol\_type="mRNA"  
/cultivar="Pomeroy"  
/db\_xref="taxon:37690"  
/clone="UCRPT02-57B01-D2-1-6.b"  
/tissue\_type="root"  
/dev\_stage="seedling"  
/lab\_host="E. coli TJC121"



```

/dev_stage="cleaving embryo"
/clone_lib="Molgula tectiformis unpublished cDNA library,
cleaving embryo"

ORIGIN
Query Match          74.4%; Score 18.6; DB 7; Length 742;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCTGTCGA 25
    ||||| ||||| ||||| |||||
Db 57 CAAATTATAGAGATTATTCTGCAA 33

RESULT 50
LOCUS BUI30653/c 745 bp mRNA linear EST 25-NOV-2002
DEFINITION 601117514F1 CSEQCHL21 Gallus gallus cDNA clone ChEST74k3 5', mRNA
sequence.
ACCESSION BUI30653
VERSION BUI30653.1 GI:25342572
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 745)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Pong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1..745
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hise"
/db_xref="taxon:9031"
/clone="ChEST74k3"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHL21"
/note="Organ: trunks; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; Modification of pBluescript
II KS(+) [Stratagene] vector to accommodate cDNA produced
with the 5'-trimmed protocol (construction of
uni-directionally cloned cDNA libraries from messenger RNA
for improved 3' end DNA sequencing by Glenn Fu, et al.
U.S. Patent # 6,387,624). Cut pBluescript II KS(+) with
NotI and EcoRI. Ligate in double stranded adaptor
containing BglI and BamHI sites
[5'-ggccgcgtgcagccgcggatccggaataaaag]
[5'-aattctttttcggatccggggtgcagc]"

ORIGIN
Query Match          74.4%; Score 18.6; DB 5; Length 745;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCTGTCGA 25
    ||||| ||||| ||||| |||||
Db 138 CAAATTAAAGAGACTGCTAGTTCAA 114

RESULT 51
LOCUS CJ358434/c 754 bp mRNA linear EST 15-JUN-2005
DEFINITION CJ358434 Molgula tectiformis unpublished cDNA library, cleaving
embryo Molgula tectiformis cDNA clone mtc1007o17 3', mRNA sequence.
ACCESSION CJ358434
VERSION CJ358434.1 GI:67798582
KEYWORDS EST.
SOURCE Molgula tectiformis
ORGANISM Molgula tectiformis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Stolidobranchia; Molgulidae; Molgula.
REFERENCE 1 (bases 1 to 754)
AUTHORS Gyoja, F., Satou, Y. and Satoh, N.
Expressed genes in Molgula tectiformis
Unpublished (2005)
JOURNAL
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
When you want to obtain this EST clone, please send an e-mail to
Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka
Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).

FEATURES
source
1..754
/organism="Molgula tectiformis"
/mol_type="mRNA"
/db_xref="taxon:30286"
/clone="mtc1007o17"
/tissue_type="whole animal"
/dev_stage="cleaving embryo"
/clone_lib="Molgula tectiformis unpublished cDNA library,
cleaving embryo"

ORIGIN
Query Match          74.4%; Score 18.6; DB 7; Length 754;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCTGTCGA 25
    ||||| ||||| ||||| |||||
Db 37 CAAATTATAGAGATTATTCTGCAA 13

RESULT 52
LOCUS BZ246552 758 bp DNA linear GSS 12-OCT-2002
DEFINITION CH230-457M24.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-457M24, genomic survey sequence.
ACCESSION BZ246552
VERSION BZ246552.1 GI:23906816
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 758)
AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsagaye, G., Geer, K.,
Shvartsbeyn, A., Gebregeorgis, E., Overton, L., Russell, D., Chen, D.,
Riggs, P., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other_GSS: CH230-457M24.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

```

Email: szhao@tigr.org  
 Clones are derived from the rat BAC library CHORI-230  
 (http://www.chori.org/bacpac/rat230.htm). For BAC library  
 availability, please contact Pieter de Jong (pdejong@mail.chori.org).  
 Clones may be purchased from BACPAC Resources  
 (http://www.chori.org/bacpac/or ering information.htm). BAC end  
 page: http://www.tigr.org/tdb/bac\_ends/rat/bac\_end\_intro.html  
 Plate: 457 row: M column: 24  
 Seq primer: SP6  
 Class: BAC ends.

## FEATURES

source  
 1. .758  
 Location/Qualifiers  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /strain="BN/SENHsd/MCW"  
 /db\_xref="taxon:10116"  
 /clone="CH230-457M24"  
 /sex="Female"  
 /cell\_type="Brain"  
 /clone\_lib="CHORI-230 Segment 2"  
 /note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;  
 CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by  
 Pieter de Jong"

## ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 758;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCAA 25  
 |||||  
 Db 8 CAAATTAAAGAGACTATTCGTGAAA 32

## RESULT 53

CV708436 764 bp mRNA linear EST 03-NOV-2004  
 LOCUS UCRPT01\_0011F05 f Poncirus trifoliata CTV-challenged cDNA library -  
 DEFINITION AG12 Poncirus trifoliata cDNA clone PT\_6Ea0011F05, mRNA sequence.

## ACCESSION

CV708436  
 EST.  
 Poncirus trifoliata  
 Poncirus trifoliata  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Sapindales; Rutaceae; Poncirus.  
 1 (bases 1 to 764)

## REFERENCE

AUTHORS  
 Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,  
 Wanmaker, S., Kim, H.R., Kudrna, D. and Stum, D., Wissotski, M.,  
 Wing, R.  
 Development of EST Resources and New Genetic Markers for California  
 Citrus - Poncirus trifoliata CTV-challenged phloem - AG12  
 Unpublished (2004)

## TITLE

JOURNAL  
 COMMENT  
 Contact: Mikeal Roose  
 Department of Botany & Plant Sciences, University of California  
 Riverside, CA, 92521-0124, USA  
 Tel: 9097874137  
 Fax: 9097874437  
 Email: mikeal.roose@ucr.edu

## FEATURES

source  
 1. .764  
 Location/Qualifiers  
 /organism="Poncirus trifoliata"  
 /mol\_type="mRNA"  
 /cultivar="Pomeroy Op"  
 /db\_xref="taxon:37690"  
 /clone="PT\_6Ea0011F05"  
 /tissue\_type="Phloem"  
 /dev\_stage="10 - 30 cm shoots"  
 /lab\_host="E. coli TJC121"  
 /clone\_lib="Poncirus trifoliata CTV-challenged cDNA  
 library - AG12"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
 Site 1: EcoRI; Site 2: XhoI; Plants were grown in the  
 greenhouse at University of California, Riverside. The  
 action was an open-pollinated (very probably selfed)  
 seedling of Poncirus trifoliata cv Pomeroy that was  
 selected as homozygous for the Ctv resistance gene. The  
 rootstock was sweet orange infected with citrus tristeza  
 virus (CTV) isolate T514 over 1 year before sampling (CTV  
 infects sweet orange, but not genotypes carrying the Ctv  
 resistance gene). Shoots 10-30 cm long were harvested in  
 October 2000, and the green phloem (bark) was removed and  
 frozen quickly in dry ice. Total RNA was extracted using  
 TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA  
 library was made, and 0.5 million primary lambda cDNA  
 clones were in vivo excised to give a population of  
 pBluescript SK(-) phagemids. All steps to this point were  
 performed in the M. Roose lab at the University of  
 California, Riverside by X. Ye. Phagemids were plated,  
 plasmid DNA purified, cDNA clones archived, and DNA  
 sequences determined bi-directionally using an ABI3730 at  
 the Arizona Genomics Institute, University of Arizona  
 (Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files  
 were downloaded to UC Riverside (Close), then processed at  
 UC Riverside (Wanmaker) using the HarVest pipeline  
 (http://harvest.ucr.edu) to remove vector and cloning  
 oligo sequences and various contaminants, and to trim to a  
 high quality region. Sequences that retained a phred 17  
 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 764;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCAA 25  
 |||||  
 Db 115 CAAATTAAATACATATTCGTGCAA 139

## RESULT 54

CV644101 788 bp mRNA linear EST 18-JAN-2005  
 LOCUS UCRPT02\_71G08 b Poncirus trifoliata Roots with Iron Deficiency -  
 DEFINITION UCRPT02-UCR1 Poncirus trifoliata cDNA clone UCRPT02-71G08-N16-6.b,  
 mRNA sequence.

## ACCESSION

CV644101  
 EST.  
 Poncirus trifoliata  
 Poncirus trifoliata  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
 rosids; eurosids II; Sapindales; Rutaceae; Poncirus.  
 1 (bases 1 to 788)

## REFERENCE

AUTHORS  
 Close, T.J., Roose, M.L., Parker, D.R., Federici, C.F., Mandal, J.,  
 Fenton, R.D., Wanmaker, S., Landry, B., Hubert, N., Laforest, M.,  
 Landry, J. and Ligonde, A.  
 Development of EST Resources and New Genetic Markers for California  
 Citrus - Poncirus trifoliata Roots with Iron Deficiency -  
 UCRPT02-UCR1  
 Unpublished (2005)

## JOURNAL

COMMENT  
 Contact: Timothy J. Close  
 Department of Botany & Plant Sciences  
 University of California  
 Riverside, CA 92521-0124, USA  
 Tel: 909-787-3318  
 Fax: 909-787-4437  
 Email: timothy.close@ucr.edu

## Seq primer: T7;

Location/Qualifiers

1. .788  
 /organism="Poncirus trifoliata"  
 /mol\_type="mRNA"



/cultivar="Pomeroy"  
 /db\_xref="taxon:37690"  
 /clone="UCRPT02-71G08-N16-6.b"  
 /tissue\_type="root"  
 /dev\_stage="seedling"  
 /lab\_host="E. coli TJC121"  
 /clone\_lib="Poncirus trifoliata Roots with Iron Deficiency - UCRPT02-UCR1"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plant materials were prepared by Federici (Roose lab) with advice from Parker. Seedlings 138 days after sowing were about 20-23 cm in height, bearing 8-12 leaves. On May 26, 2004, plants were washed free of soil with a stream of water, then placed upright with the roots submerged in two tanks of nutrient solution. The solution was sufficient in all major and minor nutrients and buffered with MES at about pH 6.95. The plants were maintained in this until June 14, 2004, when the solution was changed to one with only 20 micromolar iron, chelated with EDTA to induce iron deficiency. The pH was maintained at 6.99 by sodium carbonate/CO2 buffering. This solution was replaced on July 6, 2004. Roots from three plants were sampled on June 16, June 21, July 1 and July 28, 2004. Roots were collected by removing the plant from the nutrient solution, blotting off excess moisture with a paper towel, then cutting off the top of the plants. Three plants were pooled in one aluminum foil packet, and frozen between two sheets of dry ice. The time between removal from solution and freezing on dry ice did not exceed one minute. The frozen tissue was stored in these foil packs at -80C. This sampling strategy did not correspond to initiation of stress in the plant, but only to when the plants entered the low iron solution. The actual stress was not initiated until the internal iron reserve was depleted. By the July 1 sampling date, slight iron deficiency symptoms were apparent. By the final sampling date, clear iron deficiency symptoms were present. By the July 1 sampling date, the roots showed growth of Thielaviopsis brevicola, a fungus. Other than what came away when the roots were blotted with paper towels, no effort was made to remove the fungus because it is not just a surface contaminant, but grows within the roots as well. Mandal and Fenton (Close lab) purified RNA using TRIzol, poly(A) mRNA using an Oligotex mRNA Kit (Qiagen), produced a primary cDNA library using a lambda ZAP XR cDNA Synthesis Kit (Stratagene), then mass-excised 0.5 million pfu from the primary library to produce a phagemid population. The library was made from equal portions of RNA from each of the four collection dates. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at DNA Landmarks (Landry, Hubert, Laforest, Landry, Ligonde). Chromatogram files were downloaded by FTP by Close, then processed by Wanmaker (Close lab) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were assembled, then chimeras were removed following manual inspection of assemblies (Close, Roose, Wanmaker). Sequences that survived all removal steps were submitted to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 8; Length 788;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTCAA 25  
 |||||  
 Db 137 CAAATTAAATACAATATTCGTCAA 161  
 |||||

## RESULT 55

CV706539

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 793

Location/Qualifiers

/organism="Poncirus trifoliata"

/mol\_type="mRNA"

/cultivar="Pomeroy OP"

/db\_xref="taxon:37690"

/clone="PT 6EA0008H19"

/tissue\_type="Phloem"

/dev\_stage="10 - 30 cm shoots"

/lab\_host="E. coli TJC121"

/clone\_lib="Poncirus trifoliata CTV-challenged cDNA library - AGI2"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the greenhouse at University of California, Riverside. The scion was an open-pollinated (very probably selfed) seedling of Poncirus trifoliata cv Pomeroy that was selected as homozygous for the Ctv resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate TS14 over 1 year before sampling (CTV infects sweet orange, but not genotypes carrying the Ctv resistance gene). Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA library was made, and 0.5 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript SK(-) phagemids. All steps to this point were performed in the ML Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files were downloaded to UC Riverside (Close), then processed at UC Riverside (Wanmaker) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 793;  
 Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

CV706539 793 bp mRNA linear EST 03-NOV-2004  
 UCRPT01\_0008H19.f Poncirus trifoliata CTV-challenged cDNA library - AGI2 Poncirus trifoliata cDNA clone PT\_6EA0008H19, mRNA sequence.

CV706539 GI:55288907

EST.

Poncirus trifoliata

Poncirus trifoliata

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicotyledons; rosids; eurosids II; Sapindales; Rutaceae; Poncirus.

1 (bases 1 to 793)

Roose,M.L., Ye,X., Federici,C.F., Close,T.J., Fenton,R.D.,

Wanmaker,S., Kim,H.R., Kudrna,D. and Stum,D., Wissotski,M.,

Wing,R.

Development of EST Resources and New Genetic Markers for California

Citrus - Poncirus trifoliata CTV-challenged phloem - AGI2

Unpublished (2004)

Contact: Mikeal Roose

Department of Botany &amp; Plant Sciences, University of California

Riverside, CA, 92521-0124, USA

Tel: 9097874137

Fax: 9097874437

Email: mikeal.roose@ucr.edu

Seq primer: T7.

Location/Qualifiers

1. 793

/organism="Poncirus trifoliata"

/mol\_type="mRNA"

/cultivar="Pomeroy OP"

/db\_xref="taxon:37690"

/clone="PT 6EA0008H19"

/tissue\_type="Phloem"

/dev\_stage="10 - 30 cm shoots"

/lab\_host="E. coli TJC121"

/clone\_lib="Poncirus trifoliata CTV-challenged cDNA library - AGI2"

/note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in the greenhouse at University of California, Riverside. The scion was an open-pollinated (very probably selfed) seedling of Poncirus trifoliata cv Pomeroy that was selected as homozygous for the Ctv resistance gene. The rootstock was sweet orange infected with citrus tristeza virus (CTV) isolate TS14 over 1 year before sampling (CTV infects sweet orange, but not genotypes carrying the Ctv resistance gene). Shoots 10-30 cm long were harvested in October 2000, and the green phloem (bark) was removed and frozen quickly in dry ice. Total RNA was extracted using TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA library was made, and 0.5 million primary lambda cDNA clones were in vivo excised to give a population of pBluescript SK(-) phagemids. All steps to this point were performed in the ML Roose lab at the University of California, Riverside by X. Ye. Phagemids were plated, plasmid DNA purified, cDNA clones archived, and DNA sequences determined bi-directionally using an ABI3730 at the Arizona Genomics Institute, University of Arizona (Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files were downloaded to UC Riverside (Close), then processed at UC Riverside (Wanmaker) using the HarVEST pipeline (<http://harvest.ucr.edu>) to remove vector and cloning oligo sequences and various contaminants, and to trim to a high quality region. Sequences that retained a phred 17 region of at least 100 bases were deposited to GenBank."



```

QY 1 CAAATTAAAGAGACTATTCTGCGAA 25
    ||||| ||||| ||||| ||||| |||||
Db 179 CAAATTAAATACATATTCGTCGAA 203

RESULT 56
CV705522 798 bp mRNA linear EST 03-NOV-2004
LOCUS UCRPT01_0007A02 f Poncirus trifoliata CTV-challenged cDNA library -
DEFINITION AG12 Poncirus trifoliata cDNA clone PT_6Ea0007A02, mRNA sequence.
ACCESSION CV705522
VERSION 1
KEYWORDS EST.
SOURCE Poncirus trifoliata
ORGANISM Poncirus trifoliata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
1 (bases 1 to 798)
REFERENCE
AUTHORS Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,
Wanamaker, S., Kim, H.R., Kudrna, D. and Stum, D., Wisotski, M.,
Wing, R.
TITLE Development of EST Resources and New Genetic Markers for California
JOURNAL Citrus - Poncirus trifoliata CTV-challenged phloem - AG12
COMMENT Unpublished (2004)
Contact: Mikeal Roose
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124, USA
Tel: 9097874137
Fax: 9097874437
Email: mikeal.roose@ucr.edu
Seq primer: T7
FEATURES
source
1..798
/organism="Poncirus trifoliata"
/mol_type="mRNA"
/cultivar="Pomeroy Op"
/db_xref="taxon:37690"
/clone="PT_6Ea0007A02"
/tissue_type="Phloem"
/lab_host="E. coli TJC121"
/dev_stage="10 - 30 cm shoots"
/clone_lib="Poncirus trifoliata CTV-challenged cDNA
library - AG12"
/notes="vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse at University of California, Riverside. The
scion was a open-pollinated (very probably selfed)
seedling of Poncirus trifoliata cv Pomeroy that was
selected as homozygous for the Ctv resistance gene. The
rootstock was sweet orange infected with citrus tristeza
virus (CTV) isolate TS14 over 1 year before sampling (CTV
infects sweet orange, but not genotypes carrying the Ctv
resistance gene). Shoots 10-30 cm long were harvested in
October 2000, and the green phloem (bark) was removed and
frozen quickly in dry ice. Total RNA was extracted using
TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA
library was made, and 0.5 million primary lambda cDNA
clones were in vivo excised to give a population of
pBluescript SK(-) phagemids. All steps to this point were
performed in the ML Roose lab at the University of
California, Riverside by X. Ye. Phagemids were plated,
plasmid DNA purified, cDNA clones archived, and DNA
sequences determined bi-directionally using an ABI3730 at
the Arizona Genomics Institute, University of Arizona
(Kim, Kudrna, Stum, Wisotski, Wing). Chromatogram files
were downloaded to UC Riverside (Close), then processed at
UC Riverside (Wanamaker) using the Harvest pipeline
(http://harvest.ucr.edu) to remove vector and cloning
oligo sequences and various contaminants, and to trim to a
high quality region. Sequences that retained a phred 17
region of at least 100 bases were deposited to GenBank."
ORIGIN

```

```

Query Match 74.4%; Score 18.6; DB 7; Length 798;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGCGAA 25
    ||||| ||||| ||||| ||||| |||||
Db 184 CAAATTAAATACATATTCGTCGAA 208

RESULT 57
BZ536243/c 798 bp DNA linear GSS 16-DEC-2002
LOCUS OGALJ05TM ZM2_0.7_1.5_KB Zea mays genomic clone ZMMBMA0087B10,
DEFINITION genomic survey sequence.
ACCESSION BZ536243
VERSION BZ536243.1 GI:27083861
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 798)
REFERENCE
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OGALJ05TC
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: methylation filtered.
FEATURES
source
1..798
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZM2_0.7_1.5_KB"
/notes="vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"
ORIGIN

Query Match 74.4%; Score 18.6; DB 9; Length 798;
Best Local Similarity 84.0%; Pred. No. 1.1e+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGCGAA 25
    ||||| ||||| ||||| ||||| |||||
Db 242 CATATTAAAGGAGCATATTCTGCGAA 218

RESULT 58
CV704979 810 bp mRNA linear EST 03-NOV-2004
LOCUS UCRPT01_0006D07 f Poncirus trifoliata CTV-challenged cDNA library -
DEFINITION AG12 Poncirus trifoliata cDNA clone PT_6Ea0006D07, mRNA sequence.
ACCESSION CV704979
VERSION CV704979.1 GI:55287347
KEYWORDS EST.
SOURCE Poncirus trifoliata
ORGANISM Poncirus trifoliata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.
1 (bases 1 to 810)
REFERENCE

```

**AUTHORS** Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,  
Wanamaker, S., Kim, H.R., Kudrna, D. and Stum, D., Wissotski, M.,  
Wing, R.

**TITLE** Development of EST Resources and New Genetic Markers for California  
Citrus - Poncirus trifoliata CTV-challenged phloem - AG12

**JOURNAL** Unpublished (2004)

**COMMENT** Contact: Mikeal Roose  
Department of Botany & Plant Sciences, University of California  
Riverside, CA, 92521-0124, USA  
Tel: 9097874137  
Fax: 9097874437  
Email: mikeal.roose@ucr.edu  
Seq primer: T7

# FEATURES

source  
1..810  
/organism="Poncirus trifoliata"  
/mol\_type="mRNA"  
/cultivar="Pomeroy OP"  
/db\_xref="taxon:37690"  
/clone="PT\_6Ea006D07"  
/tissue\_type="Phloem"  
/dev\_stage="10 - 30 cm shoots"  
/lab\_host="E. coli TJC121"  
/clone\_lib="Poncirus trifoliata CTV-challenged cDNA  
library - AG12"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the  
greenhouse at University of California, Riverside. The  
action was an open-pollinated (very probably selfed)  
seedling of Poncirus trifoliata cv Pomeroy that was  
selected as homozygous for the Ctv resistance gene. The  
rootstock was sweet orange infected with citrus tristeza  
virus (CTV) isolate T514 over 1 year before sampling (CTV  
infests sweet orange, but not genotypes carrying the Ctv  
resistance gene). Shoots 10-30 cm long were harvested in  
October 2000, and the green phloem (bark) was removed and  
frozen quickly in dry ice. Total RNA was extracted using  
TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA  
library was made, and 0.5 million primary lambda cDNA  
clones were in vivo excised to give a population of  
phagescript SK(-) phagemids. All steps to this point were  
performed in the ML Roose lab at the University of  
California, Riverside by X. Ye. Phagemids were plated,  
plasmid DNA purified, cDNA clones archived, and DNA  
sequences determined bi-directionally using an ABI3730 at  
the Arizona Genomics Institute, University of Arizona  
(Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files  
were downloaded to UC Riverside (Close), then processed at  
UC Riverside (Wanamaker) using the HarVest pipeline  
(http://harvest.ucr.edu) to remove vector and cloning  
oligo sequences and various contaminants, and to trim to a  
high quality region. Sequences that retained a phred 17  
region of at least 100 bases were deposited to GenBank."

## ORGANISM

Poncirus trifoliata  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
rosids; eurosids II; Sapindales; Rutaceae; Poncirus.  
1 (bases 1 to 816)  
Roose, M.L., Ye, X., Federici, C.F., Close, T.J., Fenton, R.D.,  
Wanamaker, S., Kim, H.R., Kudrna, D. and Stum, D., Wissotski, M.,  
Wing, R.

## REFERENCE AUTHORS

**TITLE** Development of EST Resources and New Genetic Markers for California  
Citrus - Poncirus trifoliata CTV-challenged phloem - AG12

## JOURNAL COMMENT

Contact: Mikeal Roose  
Department of Botany & Plant Sciences, University of California  
Riverside, CA, 92521-0124, USA  
Tel: 9097874137  
Fax: 9097874437  
Email: mikeal.roose@ucr.edu  
Seq primer: T7

## FEATURES source

1..816  
/organism="Poncirus trifoliata"  
/mol\_type="mRNA"  
/cultivar="Pomeroy OP"  
/db\_xref="taxon:37690"  
/clone="PT\_6Ea001K18"  
/tissue\_type="Phloem"  
/dev\_stage="10 - 30 cm shoots"  
/lab\_host="E. coli TJC121"  
/clone\_lib="Poncirus trifoliata CTV-challenged cDNA  
library - AG12"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the  
greenhouse at University of California, Riverside. The  
action was an open-pollinated (very probably selfed)  
seedling of Poncirus trifoliata cv Pomeroy that was  
selected as homozygous for the Ctv resistance gene. The  
rootstock was sweet orange infected with citrus tristeza  
virus (CTV) isolate T514 over 1 year before sampling (CTV  
infests sweet orange, but not genotypes carrying the Ctv  
resistance gene). Shoots 10-30 cm long were harvested in  
October 2000, and the green phloem (bark) was removed and  
frozen quickly in dry ice. Total RNA was extracted using  
TRIzol reagent (Gibco). Poly(A) RNA was purified, a cDNA  
library was made, and 0.5 million primary lambda cDNA  
clones were in vivo excised to give a population of  
phagescript SK(-) phagemids. All steps to this point were  
performed in the ML Roose lab at the University of  
California, Riverside by X. Ye. Phagemids were plated,  
plasmid DNA purified, cDNA clones archived, and DNA  
sequences determined bi-directionally using an ABI3730 at  
the Arizona Genomics Institute, University of Arizona  
(Kim, Kudrna, Stum, Wissotski, Wing). Chromatogram files  
were downloaded to UC Riverside (Close), then processed at  
UC Riverside (Wanamaker) using the HarVest pipeline  
(http://harvest.ucr.edu) to remove vector and cloning  
oligo sequences and various contaminants, and to trim to a  
high quality region. Sequences that retained a phred 17  
region of at least 100 bases were deposited to GenBank."

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 810;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGTGCAA 25  
|||||  
DB 190 CAATTAATACAAATATCCGTGCAA 214  
|||||

## RESULT 59

CV708661 816 bp mRNA linear EST 03-NOV-2004  
LOCUS UCRPT01\_001K18 f Poncirus trifoliata CTV-challenged cDNA library -  
DEFINITION AG12 Poncirus trifoliata cDNA clone PT\_6Ea001K18, mRNA sequence.  
ACCESSION CV708661  
VERSION CV708661.1 GI:55291029  
KEYWORDS EST.  
SOURCE Poncirus trifoliata

## ORIGIN

Query Match 74.4%; Score 18.6; DB 7; Length 816;  
Best Local Similarity 84.0%; Pred. No. 1.1e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGTGCAA 25  
|||||  
DB 175 CAATTAATACAAATATCCGTGCAA 199  
|||||

## RESULT 60

CV708661/c 823 bp mRNA linear EST 21-JUN-2005  
LOCUS CJ401436 Molgula tectiformis unpublished cDNA library, gonad  
DEFINITION CJ401436

Molgula tectiformis cDNA clone mtgd015b02 3', mRNA sequence.

```

ACCESSION   CJ401436
VERSION     CJ401436.1  GI:68093391
KEYWORDS    EST.
SOURCE      Molgula tectiformis
            Molgula tectiformis
            Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
            Stolidobranchia; Molgulidae; Molgula.
REFERENCE   1 (bases 1 to 823)
AUTHORS     Gyoja,F., Satou,Y. and Satoh,N.
TITLE       Expressed genes in Molgula tectiformis
JOURNAL     Unpublished (2005)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp
            When you want to obtain this EST clone, please send an e-mail to
            Nori Satoh (satoh@ascidian.zool.kyoto-u.ac.jp) and its cc to Yutaka
            Satou (yutaka@ascidian.zool.kyoto-u.ac.jp).
FEATURES             Location/Qualifiers
     source          1..823
                     /organism="Molgula tectiformis"
                     /mol_type="mRNA"
                     /db_xref="taxon:30286"
                     /clone="mtgd015b02"
                     /tissue_type="gonad"
                     /clone_lib="Molgula tectiformis unpublished cDNA library,
                     gonad"
ORIGIN

Query Match      74.4%; Score 18.6; DB 7; Length 823;
Best Local Similarity 84.0%; Pred.No.1.le+03;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy  1  CAATTAAGAGAGACTATTCTGCAAA 25
    ||||| ||||| ||||| ||||| |||||
Db  62  CAATTAAGAGAGATTATTCTGCAAA 38

```

Search completed: January 28, 2006, 01:14:26  
Job time : 2035.12 secs

***This Page Blank (uspto)***

GenCore version 5.1.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:56:54 ; Search time 70.2247 Seconds  
(without alignments)  
632.812 Million cell updates/sec

Title: US-10-716-005-3

Perfect score: 25

Sequence: 1 caaattaaagagactattctgca 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/PPCTUS COMB.seq:\*

7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*

8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*

9: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	77.6	162841	3	US-09-949-016-13733
2	18.2	72.8	600	3	Sequence 3094, Ap
3	17.6	70.4	512	2	US-09-328-352-3094
4	17.6	70.4	577	3	Sequence 2, Appl
5	17.6	70.4	577	3	US-09-270-767-7869
6	17.6	70.4	2947	3	US-09-270-767-23151
7	17.6	70.4	3285	3	Sequence 3805, Ap
8	17.6	70.4	3730	3	US-09-710-279-3153
9	17.6	70.4	30549	3	Sequence 3499, Ap
10	17.6	70.4	30549	3	US-09-949-016-14328
11	17.6	70.4	236474	3	Sequence 14328, A
12	17.6	70.4	260247	3	US-09-134-001C-322
13	17.6	70.4	1664976	3	US-09-949-016-13418
14	17.6	70.4	1664976	3	US-09-949-016-13358
15	17.4	69.6	13201	3	US-08-916-421B-1
16	17.2	68.8	490	3	Sequence 1, Appl
17	17.6	70.4	601	3	US-09-692-570-1
18	17.6	70.4	601	3	US-09-949-016-14672
19	17.6	70.4	601	3	US-09-513-999C-32487
20	17.6	70.4	601	3	Sequence 140776, A
21	17.6	70.4	601	3	US-09-949-016-140776
22	17.6	70.4	601	3	Sequence 140777, A
23	17.6	70.4	601	3	US-09-949-016-142111
24	17.6	70.4	601	3	Sequence 653, App
25	17.6	70.4	601	3	US-08-956-171E-653
26	17.6	70.4	601	3	Sequence 553, App
27	17.6	70.4	601	3	US-08-781-986A-653
28	17.6	70.4	601	3	Sequence 1191, Ap
29	17.6	70.4	601	3	US-09-583-110-1191
30	17.6	70.4	601	3	Sequence 1469, Ap
31	17.6	70.4	601	3	US-09-107-433-1469
32	17.6	70.4	601	3	Sequence 12, Appl
33	17.6	70.4	601	3	US-10-081-923-12

25	17	68.0	1378	3	US-10-081-923-14	Sequence 14, Appl
26	17	68.0	1379	3	US-10-081-923-7	Sequence 7, Appl
27	17	68.0	1379	3	US-10-081-923-13	Sequence 13, Appl
28	17	68.0	1379	3	US-10-081-923-16	Sequence 16, Appl
29	17	68.0	1384	3	US-10-081-923-8	Sequence 8, Appl
30	17	68.0	1384	3	US-10-081-923-10	Sequence 10, Appl
31	17	68.0	1385	3	US-10-081-923-9	Sequence 9, Appl
32	17	68.0	1390	3	US-10-081-923-11	Sequence 11, Appl
33	17	68.0	1393	3	US-10-081-923-15	Sequence 15, Appl
34	17	68.0	1426	3	US-09-075-460-1	Sequence 1, Appl
35	17	68.0	1447	3	US-10-081-923-17	Sequence 17, Appl
36	17	68.0	1545	3	US-09-248-796A-46	Sequence 46, Appl
37	17	68.0	1795	2	US-08-579-667-1	Sequence 1, Appl
38	17	68.0	1803	3	US-10-081-923-5	Sequence 5, Appl
39	17	68.0	2114	3	US-10-104-047-1637	Sequence 1637, Ap
40	17	68.0	2811	3	US-09-248-796A-740	Sequence 740, App
41	17	68.0	8195	3	US-08-961-527-94	Sequence 94, Appl
42	17	68.0	44064	3	US-09-949-016-12015	Sequence 12015, A
43	17	68.0	44072	3	US-09-949-016-15757	Sequence 15757, A
44	17	68.0	88758	3	US-09-949-016-13502	Sequence 13502, A
45	17	68.0	96845	3	US-09-949-016-13658	Sequence 13658, A
46	17	68.0	264358	3	US-09-949-016-15725	Sequence 15725, A
47	16.8	67.2	346	3	US-09-325-932A-133	Sequence 133, App
48	16.8	67.2	1653	3	US-09-325-932A-132	Sequence 132, App
49	16.6	66.4	462	3	US-09-798-635A-10	Sequence 10, Appl
50	16.6	66.4	506	3	US-09-621-976-18917	Sequence 18917, A
51	16.6	66.4	526	3	US-09-270-767-4638	Sequence 4638, Ap
52	16.6	66.4	526	3	US-09-270-767-19920	Sequence 19920, A
53	16.6	66.4	601	3	US-09-949-016-34577	Sequence 34577, A
54	16.6	66.4	601	3	US-09-949-016-34578	Sequence 34578, A
55	16.6	66.4	601	3	US-09-949-016-48887	Sequence 48887, A
56	16.6	66.4	601	3	US-09-949-016-48888	Sequence 48888, A
57	16.6	66.4	601	3	US-09-949-016-49059	Sequence 49059, A
58	16.6	66.4	601	3	US-09-949-016-49060	Sequence 49060, A
59	16.6	66.4	601	3	US-09-949-016-49231	Sequence 49231, A
60	16.6	66.4	601	3	US-09-949-016-49232	Sequence 49232, A

ALIGNMENTS

RESULT 1  
US-09-949-016-13733  
; Sequence 13733, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13733  
; LENGTH: 162841  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13733

Query Match 77.6%; Score 19.4; DB 3; Length 162841;  
Best Local Similarity 95.2%; Pred. No. 30;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTG 22

Db 12094 AAATTAAGAGACTATTCGTG 12114

```

RESULT 2
US-09-328-352-3094
; Sequence 3094, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 252
; SEQ ID NO 3094
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3094

Query Match 72.8%; Score 18.2; DB 3; Length 600;
Best Local Similarity 87.0%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCA 24
    ||||| ||||| ||||| ||||| |||||
DB 83 AAATTAAGAGAACTATTTCGTGAA 105

RESULT 3
US-08-460-739-2
; Sequence 2, Application US/08460739
; Patent No. 5698416
; GENERAL INFORMATION:
; APPLICANT: Wolf, Marcia K.
; APPLICANT: Cassels, Frederick J.
; APPLICANT: Bell, Brian A.
; TITLE OF INVENTION: Improved Methods for Production of
; TITLE OF INVENTION: Antigens Under Control of Tempature-Regulated Promotors
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Glenna Hendricks
; STREET: 9669 A Main Street
; CITY: Fairfax
; STATE: VA
; COUNTRY: US
; ZIP: 22031
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,739
; FILING DATE: 20-APR-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Hendricks, Glenna M.
; REGISTRATION NUMBER: 32,535
; REFERENCE/DOCKET NUMBER: wolf2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 425-4250
; TELEFAX: (703) 425-2767
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 512 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:

```

```

; ORGANISM: DNA for encoding CFAI protein
US-08-460-739-2

Query Match 70.4%; Score 17.6; DB 2; Length 512;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCA 25
    ||||| ||||| ||||| ||||| |||||
DB 5 AAATTAAGAACTATTTCGTGCA 28

RESULT 4
US-09-270-767-7869
; Sequence 7869, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7869
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-7869

Query Match 70.4%; Score 17.6; DB 3; Length 577;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCA 25
    ||||| ||||| ||||| ||||| |||||
DB 478 AAATTAAGCAACTATTTCGCA 501

RESULT 5
US-09-270-767-23151
; Sequence 23151, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 23151
; LENGTH: 577
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-23151

Query Match 70.4%; Score 17.6; DB 3; Length 577;
Best Local Similarity 83.3%; Pred. No. 1.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCA 25
    ||||| ||||| ||||| ||||| |||||
DB 478 AAATTAAGCAACTATTTCGCA 501

RESULT 6
US-09-710-279-3805
; Sequence 3805, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

```

; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3805  
; LENGTH: 2947  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence  
US-09-710-279-3805

Query Match 70.4%; Score 17.6; DB 3; Length 2947;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCA 24  
Db 1894 CAAATAAAGAGAAATAAGTAGTCA 1917

RESULT 7  
US-09-710-279-3153  
; Sequence 3153, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3153  
; LENGTH: 3285  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence  
US-09-710-279-3153

Query Match 70.4%; Score 17.6; DB 3; Length 3285;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCA 24  
Db 1447 CAAATAAAGAGAAATAAGTAGTCA 1470

RESULT 8  
US-09-710-279-3499  
; Sequence 3499, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3499  
; LENGTH: 3730

; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence  
US-09-710-279-3499

Query Match 70.4%; Score 17.6; DB 3; Length 3730;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCA 24  
Db 1892 CAAATAAAGAGAAATAAGTAGTCA 1915

RESULT 9  
US-09-949-016-14328  
; Sequence 14328, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14328  
; LENGTH: 5418  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14328

Query Match 70.4%; Score 17.6; DB 3; Length 5418;  
Best Local Similarity 83.3%; Pred. No. 1.4e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTTCGTGCAA 25  
Db 4733 AAATTAAAGAGATTATTCTAGCAA 4756

RESULT 10  
US-09-134-001C-322  
; Sequence 322, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 322  
; LENGTH: 30549  
; TYPE: DNA  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-322

Query Match 70.4%; Score 17.6; DB 3; Length 30549;

Best Local Similarity 83.3%; Pred. No. 1.7e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCGTGCA 24  
|||||  
Db 12031 CAAATTAAGAGAAATAATAGTCA 12054

RESULT 11  
US-09-949-016-13418/c  
; Sequence 13418, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13418  
; LENGTH: 236474  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(236474)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-13418

Query Match 70.4%; Score 17.6; DB 3; Length 236474;  
Best Local Similarity 83.3%; Pred. No. 2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGCA 25  
|||||  
Db 197388 AAATTAAGAGACTATTCGTGTA 197365

RESULT 12  
US-09-949-016-13358  
; Sequence 13358, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13358  
; LENGTH: 260247  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13358

Query Match 70.4%; Score 17.6; DB 3; Length 260247;  
Best Local Similarity 83.3%; Pred. No. 2e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGCA 25  
|||||  
Db 152084 AACTTAAGAGATATGTGCAA 152107

RESULT 13  
US-08-916-421B-1  
; Sequence 1, Application US/08916421B  
; Patent No. 6503729  
; GENERAL INFORMATION:  
; APPLICANT: Bult et al.  
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus  
; Jannaschii  
; Patent No. 6503729  
; FILE REFERENCE: PB275  
; CURRENT APPLICATION NUMBER: US/08/916,421B  
; CURRENT FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: US 60/024,428  
; PRIOR FILING DATE: 1996-08-22  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 1664976  
; TYPE: DNA  
; ORGANISM: Methanococcus jannaschii  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (28222)..(28222)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (28257)..(28258)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84773)..(84773)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84808)..(84808)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (84812)..(84812)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98120)..(98120)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98159)..(98159)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98239)..(98239)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98266)..(98266)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (98343)..(98343)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (103998)..(103998)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (148948)..(148948)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (163385)..(163385)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (191989)..(191989)  
; OTHER INFORMATION: n equals a, t, c, or g  
; NAME/KEY: misc feature  
; LOCATION: (191995)..(191995)  
; OTHER INFORMATION: n equals a, t, c, or g



```
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
```

```
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
```

```
Query Match 70.4%; Score 17.6; DB 3; Length 1664976;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 2 AAATTAAAGAGACTATTTCGTGCAA 25
Db 1522836 AAATTAAAGAGAAAAATTAATGCAA 1522859
```

```
RESULT 14
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6797466
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; CURRENT FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; PRIOR FILING DATE: 1997-08-22
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
```

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (871619)..(871619)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1084830)..(1084830)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1096846)..(1096846)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1119881)..(1119881)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1130881)..(1130881)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1310988)..(1310988)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1313224)..(1313224)  
OTHER INFORMATION: n equals a, t, c, or g  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1349473)..(1349473)  
OTHER INFORMATION: n equals a, t, c, or g

Query Match 70.4%; Score 17.6; DB 3; Length 1664976;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTCGAA 25  
DB 1522836 AAATTAAGAGAAATTAATGTCAA 1522859

RESULT 15  
US-09-949-016-14672/c  
Sequence 14672, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14672  
LENGTH: 13201  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-14672

Query Match 69.6%; Score 17.4; DB 3; Length 13201;  
Best Local Similarity 94.7%; Pred. No. 1.9e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCG 20  
DB 1522836 AAATTAAGAGAAATTAATGTCAA 1522859

DB 5301 AAATTAAGAGACTATTCG 5283

RESULT 16  
US-09-513-999C-32487  
Sequence 32487, Application US/09513999C  
Patent No. 6783961  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Duclert, A.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
Patent No. 6783961  
FILE REFERENCE: 59.US2.REG  
CURRENT APPLICATION NUMBER: US/09/513,999C  
CURRENT FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/122,487  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 32487  
LENGTH: 490  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-513-999C-32487

Query Match 68.8%; Score 17.2; DB 3; Length 490;  
Best Local Similarity 86.4%; Pred. No. 1.6e+02;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCGTCGAA 25  
DB 377 ATTAAGAGACTATTAATGTCAA 398

RESULT 17  
US-09-949-016-140776  
Sequence 140776, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 140776  
LENGTH: 601  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-140776

Query Match 68.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTCGAA 25  
DB 483 CACATTAAGAGAACTATTAAGCA 507

RESULT 18  
US-09-949-016-140777  
Sequence 140777, Application US/09949016  
Patent No. 6812339

;; GENERAL INFORMATION:  
;; APPLICANT: VENTER, J. Craig et al.  
;; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
;; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
;; FILE REFERENCE: CL001307  
;; CURRENT APPLICATION NUMBER: US/09/949,016  
;; CURRENT FILING DATE: 2000-04-14  
;; PRIOR APPLICATION NUMBER: 60/241,755  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/237,768  
;; PRIOR FILING DATE: 2000-10-03  
;; PRIOR APPLICATION NUMBER: 60/231,498  
;; PRIOR FILING DATE: 2000-09-08  
;; NUMBER OF SEQ ID NOS: 207012  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 140777  
;; LENGTH: 601  
;; TYPE: DNA  
;; ORGANISM: Human  
US-09-949-016-140777

Query Match 68.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 CAATTAAAGAGACTATTCGTGCAA 25  
Db 491 CACATTAAAGAACTATTGAAGCAA 515

RESULT 19  
US-09-949-016-142111/c  
; Sequence 142111, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 142111  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-142111

Query Match 68.0%; Score 17; DB 3; Length 601;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 CAATTAAAGAGACTATTCGTGCAA 25  
Db 125 CAATTATGAGCAATTCATGCAA 101

RESULT 20  
US-08-956-171E-653/c  
; Sequence 653, Application US/08956171E  
; Patent No. 6593114  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; Gil H. Choi  
; Patrick S. Dillon  
; Craig A. Rosen

;; Steven C. Barash  
;; Michael R. Fannon  
;; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
;; NUMBER OF SEQUENCES: 5256  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Human Genome Sciences, Inc.  
;; STREET: 9410 Key West Avenue  
;; CITY: Rockville  
;; STATE: Maryland  
;; COUNTRY: USA  
;; ZIP: 20850  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
;; COMPUTER: HP Vectra 486/33  
;; OPERATING SYSTEM: MSDOS version 6.2  
;; SOFTWARE: ASCII Text  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/956,171E  
;; FILING DATE: 20-Oct-1997  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 60/009,861  
;; FILING DATE: January 5, 1996  
;; APPLICATION NUMBER: 08/781,986  
;; FILING DATE: January 3, 1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Mark J. Hyman  
;; REGISTRATION NUMBER: 46,789  
;; REFERENCE/DOCKET NUMBER: PB248P1  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (240) 314-1224  
;; TELEFAX: (301) 309-8439  
;; INFORMATION FOR SEQ ID NO: 653:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 747 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: double  
;; TOPOLOGY: linear  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 653:  
US-08-956-171E-653  
Query Match 68.0%; Score 17; DB 3; Length 747;  
Best Local Similarity 80.0%; Pred. No. 2.1e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 CAATTAAAGAGACTATTCGTGCAA 25  
Db 296 CAAATCAATGACAAAATTCGTGCAA 272  
RESULT 21  
US-08-781-986A-653/c  
; Sequence 653, Application US/08781986A  
; Patent No. 6737248  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 5255  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/781,986A  
; FILING DATE:

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 653:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 747 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-781-986A-653

Query Match 68.0%; Score 17; DB 3; Length 747;
Best Local Similarity 80.0%; Pred. No. 2.1e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTGCAA 25
Db 296 CAAATCAATGACAAATTCGTGCAA 272

RESULT 22
US-09-583-110-1191
; Sequence 1191, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT APPLICATION NUMBER: US/09/583,110
; CURRENT FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/107,433
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/085,131
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: US 60/051,553
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 1191
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; US-09-583-110-1191

Query Match 68.0%; Score 17; DB 3; Length 1140;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTGCAA 25
Db 265 CAAATCAAGGAAACTATCCGTGCGA 289

RESULT 23
US-09-107-433-1469
; Sequence 1469, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
```

```
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 1469:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1287 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1287
; SEQUENCE DESCRIPTION: SEQ ID NO: 1469:
; US-09-107-433-1469

Query Match 68.0%; Score 17; DB 3; Length 1287;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTGCAA 25
Db 319 CAAATCAAGGAAACTATCCGTGCGA 343

RESULT 24
US-10-081-923-12
; Sequence 12, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: pti1 sequence from isolate no. 2
; US-10-081-923-12

Query Match 68.0%; Score 17; DB 3; Length 1323;
```

US-10-081-923-13  
; Sequence 13, Application US/10081923  
; Patent No. 6593093  
; GENERAL INFORMATION:

## RESULT 32

```
; SEQ ID NO 1
; LENGTH: 1426
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (281)...(1201)
US-09-075-460-1

Query Match      68.0%; Score 17; DB 3; Length 1426;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTTCGTGCAA 25
   |||||
Db 455 CAGATTAAGAGAGATATTCTCGAAA 479

RESULT 35
US-10-081-923-17
; Sequence 17, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptA1 sequence from isolate no. 11
US-10-081-923-17

Query Match      68.0%; Score 17; DB 3; Length 1447;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTTCGTGCAA 25
   |||||
Db 98 CAGGTTAAAGAAACGATTTCGCGCAA 122

RESULT 36
US-09-248-796A-46
; Sequence 46, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 46
; LENGTH: 1545
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-46

Query Match      68.0%; Score 17; DB 3; Length 1545;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTTCGTGCAA 25
   |||||
Db 57 CAAATTAAGTGGCTATTACTTGAAA 81

RESULT 37
US-08-579-667-1
; Sequence 1, Application US/08579667
; Patent No. 5705624
; GENERAL INFORMATION:
; APPLICANT: Fitzmaurice, Wayne P.
; APPLICANT: Hellmann, Gary M.
; APPLICANT: Grill, Laurence K.
; APPLICANT: Kumagai, Monto H.
; APPLICANT: Della-Cioppa, Guy R.
; TITLE OF INVENTION: DNA SEQUENCES ENCODING ENZYMES USEFUL IN PHYTOENE BIOSYNTHESIS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia C. Bennett
; STREET: 1211 East Morehead Street, PO Drawer 34009
; CITY: Charlotte
; STATE: No. 5705624th Carolina
; COUNTRY: USA
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,667
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 627-196
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1795 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 362..1591
US-08-579-667-1

Query Match      68.0%; Score 17; DB 2; Length 1795;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGAGACTATTTCGTGCAA 25
   |||||
Db 57 CAAATTAAGTGGCTATTACTTGAAA 81

RESULT 38
US-10-081-923-5
; Sequence 5, Application US/10081923
; Patent No. 6593093
; GENERAL INFORMATION:
; APPLICANT: Uhl, James R.
; APPLICANT: Cockerill, Franklin R.
; TITLE OF INVENTION: Detection of Group A Streptococcus
; FILE REFERENCE: 07039-306001
; CURRENT APPLICATION NUMBER: US/10/081,923
```



```
; CURRENT FILING DATE: 2002-07-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Group A Streptococcus
; FEATURE:
; OTHER INFORMATION: ptai sequence from Oklahoma University M1 strain
; PUBLICATION INFORMATION:
; AUTHORS: Ferretti et al.
; JOURNAL: Proc. Natl. Acad. Sci. USA
; VOLUME: 98
; PAGES: 4658-4663
; DATE: 2001-01-01
; US-10-081-923-5

Query Match      68.0%; Score 17; DB 3; Length 1803;
Best Local Similarity 80.0%; Pred. No. 2.3e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25
Db 265 CAGTTAAAGAAACGATTCGCGCAA 289

RESULT 39
US-10-104-047-1637
; Sequence 1637, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No 6943241el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1637
; LENGTH: 2114
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-104-047-1637

Query Match      68.0%; Score 17; DB 3; Length 2114;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25
Db 1172 CAAATTAAAGAAATATTGTGTAA 1196

RESULT 40
US-09-248-796A-740
; Sequence 740, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 740
; LENGTH: 2811

; TYPE: DNA
; ORGANISM: Candida albicans
; US-09-248-796A-740

Query Match      68.0%; Score 17; DB 3; Length 2811;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25
Db 190 CAAATTATAGTGACTACTCTCTGAAA 214

RESULT 41
US-08-961-527-94
; Sequence 94, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-94

Query Match      68.0%; Score 17; DB 3; Length 8195;
Best Local Similarity 80.0%; Pred. No. 2.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCAA 25
Db 1272 CAAATCAAGGAAACTATCCGTGCGA 1296

RESULT 42
US-09-949-016-12015
; Sequence 12015, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
```

; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12015  
; LENGTH: 44064  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12015

Query Match 68.0%; Score 17; DB 3; Length 44064;  
Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 26935 CAAATTCATGAGCAATTCATGCAA 26959

## RESULT 43

US-09-949-016-15757  
; Sequence 15757, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15757  
; LENGTH: 44072  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15757

Query Match 68.0%; Score 17; DB 3; Length 44072;  
Best Local Similarity 80.0%; Pred. No. 3.2e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 26935 CAAATTCATGAGCAATTCATGCAA 26959

## RESULT 44

US-09-949-016-13502  
; Sequence 13502, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768

; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13502  
; LENGTH: 88758  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13502

Query Match 68.0%; Score 17; DB 3; Length 88758;  
Best Local Similarity 80.0%; Pred. No. 3.4e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 6218 CACCTTAAACAGACTTTTCGTGAAA 6242

## RESULT 45

US-09-949-016-13658/c  
; Sequence 13658, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13658  
; LENGTH: 96845  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13658

Query Match 68.0%; Score 17; DB 3; Length 96845;  
Best Local Similarity 80.0%; Pred. No. 3.5e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCGTGCAA 25  
||||| ||| ||| ||| ||| ||| |||  
Db 36152 CAAATTAAAGAGACTTTTATGTAA 36128

## RESULT 46

US-09-949-016-15725/c  
; Sequence 15725, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15725  
; LENGTH: 264358  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(264358)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15725

Query Match 68.0%; Score 17; DB 3; Length 264358;  
Best Local Similarity 80.0%; Pred. No. 3.8e+02;  
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGCGAA 25  
Db 167433 CACATTAAAGAACTATTGAGCAA 167409

## RESULT 47

US-09-325-932A-133  
; Sequence 133, Application US/09325932A

; Patent No. 6451604

; GENERAL INFORMATION:

; APPLICANT: Flinn, Barry

; APPLICANT: Lasham, Annette

; TITLE OF INVENTION: Compositions affecting programmed cell

; TITLE OF INVENTION: death and their use in the modification of forestry plant develop

; FILE REFERENCE: 1022

; CURRENT APPLICATION NUMBER: US/09/325,932A

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 206

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 133

; LENGTH: 346

; TYPE: DNA

; ORGANISM: Pinus radiata

US-09-325-932A-133

## Query Match

Best Local Similarity 67.2%; Score 16.8; DB 3; Length 346;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTG 20  
Db 89 CAAATCAAGAGGCTATTCTG 108

## RESULT 48

US-09-325-932A-132

; Sequence 132, Application US/09325932A

; Patent No. 6451604

; GENERAL INFORMATION:

; APPLICANT: Flinn, Barry

; APPLICANT: Lasham, Annette

; TITLE OF INVENTION: Compositions affecting programmed cell

; TITLE OF INVENTION: death and their use in the modification of forestry plant develop

; FILE REFERENCE: 1022

; CURRENT APPLICATION NUMBER: US/09/325,932A

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 206

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 132

; LENGTH: 1653

; TYPE: DNA

; ORGANISM: Pinus radiata

US-09-325-932A-132

## Query Match

Best Local Similarity 67.2%; Score 16.8; DB 3; Length 1653;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTG 20

Db 335 CAAATCAAGAGGCTATTCTG 354

## RESULT 49

US-09-798-635A-10

; Sequence 10, Application US/09798635A

; Patent No. 6936432

; GENERAL INFORMATION:

; APPLICANT: Gopalan, Venkat

; APPLICANT: Jovanovic, Milan

; APPLICANT: Eder, Paul S.

; APPLICANT: Giordano, Tony

; APPLICANT: Powers, Gordon D.

; APPLICANT: Xavier, K. Ashish

; TITLE OF INVENTION: No. 6936432el Bacterial RNase P Proteins and

; TITLE OF INVENTION: Their Use in Identifying Antibacterial Compounds

; FILE REFERENCE: 50093/016002

; CURRENT APPLICATION NUMBER: US/09/798,635A

; CURRENT FILING DATE: 2001-03-01

; PRIOR APPLICATION NUMBER: 09/516.061

; NUMBER OF SEQ ID NOS: 113

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10

; LENGTH: 462

; TYPE: DNA

; ORGANISM: Streptococcus pyogenes

US-09-798-635A-10

## Query Match 66.4%; Score 16.6; DB 3; Length 462;

Best Local Similarity 82.6%; Pred. No. 3e+02;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTCTGCGAA 25

Db 36 AATTAAGAGGACTATTCTTGCAA 58

## RESULT 50

US-09-621-976-18917

; Sequence 18917, Application US/09621976

; Patent No. 6639063

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 18917

; LENGTH: 506

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-621-976-18917

## Query Match

Best Local Similarity 66.4%; Score 16.6; DB 3; Length 506;

Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AATTAAGAGACTATTCTGCGA 24

Db 131 AATTCGAAAGACTATTCTGCA 153

## RESULT 51

US-09-270-767-4638/c

; Sequence 4638, Application US/09270767

; Patent No. 6703491

; GENERAL INFORMATION:

```
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4638
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-4638

Query Match      66.4%; Score 16.6; DB 3; Length 526;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  2 AAATTAAGAGACTATTCTGCA 24
    ||||| ||||| ||||| |||||
Db  276 AAATTAACAACAATCTTGCA 254
    ||||| ||||| ||||| |||||

RESULT 52
US-09-270-767-19920/c
; Sequence 19920, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19920
; LENGTH: 526
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-19920

Query Match      66.4%; Score 16.6; DB 3; Length 526;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  2 AAATTAAGAGACTATTCTGCA 24
    ||||| ||||| ||||| |||||
Db  276 AAATTAACAACAATCTTGCA 254
    ||||| ||||| ||||| |||||

RESULT 53
US-09-949-016-34577
; Sequence 34577, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34577
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-34577

Query Match      66.4%; Score 16.6; DB 3; Length 526;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  2 AAATTAAGAGACTATTCTGCA 24
    ||||| ||||| ||||| |||||
Db  276 AAATTAACAACAATCTTGCA 254
    ||||| ||||| ||||| |||||

RESULT 54
US-09-949-016-34578
; Sequence 34578, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34578
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-34578

Query Match      66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY  2 AAATTAAGAGACTATTCTGCA 24
    ||||| ||||| ||||| |||||
Db  177 AAATTAAGATAGTATTATGCA 199
    ||||| ||||| ||||| |||||

RESULT 55
US-09-949-016-48887/c
; Sequence 48887, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48887
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-48887

Query Match      66.4%; Score 16.6; DB 3; Length 601;
Best Local Similarity 82.6%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
||||| ||||| |||||  
Db 300 CAAATTACAGAGACTGCCCGTGC 278

## RESULT 56

US-09-949-016-48888/c  
; Sequence 48888, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 48888  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-48888

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
||||| ||||| |||||  
Db 278 CAAATTACAGAGACTGCCCGTGC 256

## RESULT 57

US-09-949-016-49059/c  
; Sequence 49059, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49059  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-49059

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
||||| ||||| |||||  
Db 300 CAAATTACAGAGACTGCCCGTGC 278

## RESULT 58

US-09-949-016-49060/c  
; Sequence 49060, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49060  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-49060

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
||||| ||||| |||||  
Db 278 CAAATTACAGAGACTGCCCGTGC 256

## RESULT 59

US-09-949-016-49231/c  
; Sequence 49231, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49231  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-49231

Query Match 66.4%; Score 16.6; DB 3; Length 601;  
Best Local Similarity 82.6%; Pred. No. 3.1e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCTGTC 23  
||||| ||||| |||||  
Db 300 CAAATTACAGAGACTGCCCGTGC 278

## RESULT 60

US-09-949-016-49232/c  
; Sequence 49232, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:01:27 ; Search time 433.708 Seconds  
(without alignments)  
476.668 Million cell updates/sec

Title: US-10-716-005-3

Perfect score: 25

Sequence: 1 caaataaagagactattcgtagc 25

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database : Published Applications NA Main:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:
- 3: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq:
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq:
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	25	9	US-10-716-005-3
2	20.2	80.8	525	7	US-10-282-122A-34808
3	20.2	80.8	1731	7	US-10-282-122A-36131
4	18.8	75.2	1417	7	US-10-425-114-20074
5	18.6	74.4	651	5	US-10-027-632-201516
6	18.6	74.4	651	6	US-10-027-632-201516
7	18.6	74.4	188153	9	US-10-756-149-1933
8	18.2	72.8	1482	4	US-09-925-065A-551673
9	17.8	71.2	1344	6	US-10-369-493-40385
10	17.8	71.2	1344	7	US-10-282-122A-10106
11	17.6	70.4	269	8	US-10-719-956-210524
12	17.6	70.4	271	7	US-10-437-963-5617
13	17.6	70.4	393	4	US-09-925-065A-515562
14	17.6	70.4	424	7	US-10-242-535A-39738
15	17.6	70.4	424	7	US-10-085-783A-39738
16	17.6	70.4	424	3	US-09-91-936-1224
17	17.6	70.4	427	9	US-10-978-245-1224
18	17.6	70.4	427	9	US-10-644-659A-7
19	17.6	70.4	514	4	US-09-925-065A-638093
20	17.6	70.4	514	4	US-09-925-065A-319988
21	17.6	70.4	517	4	US-09-925-065A-394501
22	17.6	70.4	537	4	US-09-925-065A-394501
23	17.6	70.4	586	4	US-09-925-065A-319987

ALIGNMENTS

RESULT 1  
US-10-716-005-3  
; Sequence 3, Application US/10716005  
; Publication No. US20050106578A1  
; GENERAL INFORMATION:  
; APPLICANT: Uhl, James R.  
; APPLICANT: Cockerill III, Franklin R.  
; APPLICANT: Aichinger, Christian  
; APPLICANT: Reiser, Astrid  
; TITLE OF INVENTION: Detection of Group B Streptococcus  
; FILE REFERENCE: 07039/460001  
; CURRENT APPLICATION NUMBER: US/10/716,005  
; CURRENT FILING DATE: 2003-11-18  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 25  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Oligonucleotide  
US-10-716-005-3

Query Match 100.0%; Score 25; DB 9; Length 25;  
Best Local Similarity 100.0%; Pred. No. 0.55;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCTGCA 25

Db 1 CAAATTAAGAGACTATTCTGCA 25

Sequence 203129, A  
Sequence 61594, A  
Sequence 61595, A  
Sequence 891228, A  
Sequence 865985, A  
Sequence 8199, Ap  
Sequence 21294, A  
Sequence 3797, Ap  
Sequence 3484, A  
Sequence 57374, A  
Sequence 4939, Ap  
Sequence 34396, A  
Sequence 326, App  
Sequence 14998, A  
Sequence 355, App  
Sequence 1282, Ap  
Sequence 83, Appl  
Sequence 122, App  
Sequence 130, App  
Sequence 9592, Ap  
Sequence 58172, A  
Sequence 114621, A  
Sequence 114622, A  
Sequence 114623, A  
Sequence 114622, A  
Sequence 114623, A  
Sequence 35, Appl  
Sequence 18587, A  
Sequence 507, App  
Sequence 11, Appl  
Sequence 2308, Ap  
Sequence 9748, Ap  
Sequence 2308, Ap  
Sequence 1810, Ap  
Sequence 988, Ap

RESULT 2  
US-10-282-122A-34808  
; Sequence 34808, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:

; APPLICANT: Wang, Liangu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 34808

; LENGTH: 525

; TYPE: DNA

; ORGANISM: Staphylococcus epidermidis

US-10-282-122A-34808

Query Match 80.8%; Score 20.2; DB 7; Length 525;

Best Local Similarity 88.0%; Pred. No. 1e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCGCA 25

||||| ||||| ||||| ||||| |||||

DB 28 CAAATTAATGAGAAATTCGTGCA 52

RESULT 3

US-10-282-122A-36131

; Sequence 36131, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; CURRENT FILING DATE: 2003-02-20

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; PRIOR APPLICATION NUMBER: 60/230,347

; PRIOR FILING DATE: 2000-09-09

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/267,636

; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 36131

; LENGTH: 1731

; TYPE: DNA

; ORGANISM: Streptococcus mutans

US-10-282-122A-36131

Query Match 80.8%; Score 20.2; DB 7; Length 1731;

Best Local Similarity 88.0%; Pred. No. 1.2e+02;

Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCTGTCGCA 25

||||| ||||| ||||| ||||| |||||

DB 265 CAGATTAAGAGACAATTCGTACAA 289

RESULT 4

US-10-425-114-20074

; Sequence 20074, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 20074

; LENGTH: 1417

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: LIB3170-074-E8\_FLI

US-10-425-114-20074

Query Match 75.2%; Score 18.8; DB 7; Length 1417;

Best Local Similarity 90.9%; Pred. No. 4.9e+02;



Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCGTCAA 25  
Db 926 ATTAAGAACTATTCGTGTA 947

RESULT 5

US-10-027-632-201516/c  
; Sequence 201516, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 201516  
; LENGTH: 651  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-201516

Query Match 74.4%; Score 18.6; DB 5; Length 651;  
Best Local Similarity 84.0%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTCAA 25  
Db 544 CAAATTAAGAGATTATTTAGCAA 520

RESULT 6

US-10-027-632-201516/c  
; Sequence 201516, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002

; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 201516  
; LENGTH: 651  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-201516

Query Match 74.4%; Score 18.6; DB 6; Length 651;  
Best Local Similarity 84.0%; Pred. No. 5.3e+02;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CAAATTAAGAGACTATTCGTCAA 25  
Db 544 CAAATTAAGAGATTATTTAGCAA 520

RESULT 7

US-10-756-149-1933/c  
; Sequence 1933, Application US/10756149  
; Publication No. US20050181375A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF METASTATIC CANCER, COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS OF SCREENING FOR MODULATORS OF METASTATIC CANCER  
; FILE REFERENCE: file  
; CURRENT APPLICATION NUMBER: US/10/756,149  
; CURRENT FILING DATE: 2004-01-12  
; NUMBER OF SEQ ID NOS: 5818  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1933  
; LENGTH: 168153  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
US-10-756-149-1933

Query Match 74.4%; Score 18.6; DB 9; Length 168153;  
Best Local Similarity 84.0%; Pred. No. 1.2e+03;  
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTCGTCAA 25  
Db 160896 CAAATTAAGAGATTATTTAGCAA 160872

RESULT 8

US-09-925-065A-551673/c  
; Sequence 551673, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 551673  
; LENGTH: 1482  
; TYPE: DNA

```

; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10106
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Bacillus anthracis
;
US-10-282-122A-10106

Query Match 71.2%; Score 17.8; DB 7; Length 1344;
Best Local Similarity 90.5%; Pred. No. 1.1e+03;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTG 22
      |||||
DB 1166 AAATTAAGAAATATTCGTG 1186

RESULT 11
US-10-719-956-210524/c
; Sequence 210524, Application US/10719956
; Publication No. US20040146910A1
; GENERAL INFORMATION:
; APPLICANT: Xue Mei Zhou
; TITLE OF INVENTION: Methods of Genetic Analysis of Rat
; FILE REFERENCE: 3527.1
; CURRENT APPLICATION NUMBER: US/10719,956
; CURRENT FILING DATE: 2003-11-20
; PRIOR APPLICATION NUMBER: 60/427,836
; PRIOR FILING DATE: 2002-11-20
; NUMBER OF SEQ ID NOS: 699466
; SOFTWARE: Microarray Probe Sequence Listing Generator V 1.1
; SEQ ID NO 210524
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Rattus norvegicus
;
US-10-719-956-210524

Query Match 70.4%; Score 17.6; DB 7; Length 25;
Best Local Similarity 83.3%; Pred. No. 9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
      |||||
DB 25 AAATTAAGGACGATTGGAGCAA 2

RESULT 12
US-10-425-115-34598
; Sequence 34598, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115

```

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 34598
; LENGTH: 269
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(269)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_131586C.1
US-10-425-115-34598
```

```
Query Match 70.4%; Score 17.6; DB 8; Length 269;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 AAATTAAGAGAGACTATTCGTGCAA 25
Db 161 AACTTAAGAGAGACTATTACTTCAA 184
```

```
RESULT 13
US-10-437-963-5617
; Sequence 5617, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 5617
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(271)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_12387C.1
US-10-437-963-5617
```

```
Query Match 70.4%; Score 17.6; DB 7; Length 271;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 AAATTAAGAGAGACTATTCGTGCAA 25
Db 46 AAATTAAGTTAATATTCTTGCAA 69
```

```
RESULT 14
US-09-925-065A-515562
; Sequence 515562, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925.065A
```

```
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 515562
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-515562
```

```
Query Match 70.4%; Score 17.6; DB 4; Length 393;
Best Local Similarity 83.3%; Pred. No. 1.3e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 AAATTAAGAGAGACTATTCGTGCAA 25
Db 177 AAATTAAGAGAGACTATTACGCAA 200
```

```
RESULT 15
US-10-242-535A-39738
; Sequence 39738, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liew, C. C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 56994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39738
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (22)..(22)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (52)..(52)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (72)..(72)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (176)..(176)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (402)..(402)
; OTHER INFORMATION: n is a, c, g, or t
```





; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 319987  
; LENGTH: 586  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-319987

Query Match 70.4%; Score 17.6; DB 4; Length 586;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 475 AGATTAAAGAGACTTTAAAGTGCAA 452

RESULT 24  
US-09-925-065A-203129  
; Sequence 203129, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 203129  
; LENGTH: 589  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-203129

Query Match 70.4%; Score 17.6; DB 4; Length 589;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 560 AAATTAAGAGATGATACCTGCAA 583

RESULT 25  
US-10-972-079-61594  
; Sequence 61594, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 61595  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken  
US-10-972-079-61595

Query Match 70.4%; Score 17.6; DB 9; Length 600;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 191 AAATTAAGAAACACTATTAGTGAAA 214

RESULT 27  
US-09-925-065A-891228/c  
; Sequence 891228, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 319987  
; LENGTH: 586  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-319987

; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DeNISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 61594  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken  
US-10-972-079-61594

Query Match 70.4%; Score 17.6; DB 9; Length 600;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 266 AAATTAAGAAACACTATTAGTGAAA 289

RESULT 26  
US-10-972-079-61595  
; Sequence 61595, Application US/10972079  
; Publication No. US20050153317A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DeNISE, Sue K.  
; APPLICANT: ROSENFELD, David  
; APPLICANT: KERR, Richard  
; APPLICANT: BATES, Stephen  
; APPLICANT: HOLM, Tom  
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEEF  
; FILE REFERENCE: MM1110-2  
; CURRENT APPLICATION NUMBER: US/10/972,079  
; CURRENT FILING DATE: 2004-10-22  
; PRIOR APPLICATION NUMBER: US 60/514,333  
; PRIOR FILING DATE: 2003-10-24  
; NUMBER OF SEQ ID NOS: 96631  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 61595  
; LENGTH: 600  
; TYPE: DNA  
; ORGANISM: Chicken  
US-10-972-079-61595

Query Match 70.4%; Score 17.6; DB 9; Length 600;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCAA 25  
| | | | | | | | | | | | | | | | | | | | | |  
Db 191 AAATTAAGAAACACTATTAGTGAAA 214

RESULT 27  
US-09-925-065A-891228/c  
; Sequence 891228, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 319987  
; LENGTH: 586  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-319987

FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925,065A  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243,096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252,147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250,092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261,766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289,846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 891228  
LENGTH: 622  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-891228

Query Match 70.4%; Score 17.6; DB 4; Length 622;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTCAA 25  
|||||  
Db 433 AAATCAACAGACTATTCATGAA 410  
|||||

RESULT 28  
US-09-925-065A-865985/c  
Sequence 865985, Application US/09925065A  
Publication No. US20050228172A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.135  
CURRENT APPLICATION NUMBER: US/09/925,065A  
PRIOR FILING DATE: 2001-08-08  
PRIOR APPLICATION NUMBER: US 60/243,096  
PRIOR FILING DATE: 2000-10-24  
PRIOR APPLICATION NUMBER: US 60/252,147  
PRIOR FILING DATE: 2000-11-20  
PRIOR APPLICATION NUMBER: US 60/250,092  
PRIOR FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/261,766  
PRIOR FILING DATE: 2001-01-16  
PRIOR APPLICATION NUMBER: US 60/289,846  
PRIOR FILING DATE: 2001-05-09  
NUMBER OF SEQ ID NOS: 957086  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 865985  
LENGTH: 627  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-065A-865985

Query Match 70.4%; Score 17.6; DB 4; Length 627;  
Best Local Similarity 83.3%; Pred. No. 1.4e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTCAA 25  
|||||  
Db 26 AAATTAAGAGATGATACCTGCAA 3  
|||||

RESULT 29  
US-10-767-701-8199  
Sequence 8199, Application US/10767701  
Publication No. US20040172684A1  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
FILE REFERENCE: 38-21(53377)B  
CURRENT APPLICATION NUMBER: US/10/739,930  
CURRENT FILING DATE: 2003-12-18

APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With Plants and Uses Thereof For Plant Improvement  
FILE REFERENCE: 38-21(53335)B  
CURRENT APPLICATION NUMBER: US/10/767,701  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 8199  
LENGTH: 916  
TYPE: DNA  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS29491\_1  
US-10-767-701-8199

Query Match 70.4%; Score 17.6; DB 7; Length 916;  
Best Local Similarity 83.3%; Pred. No. 1.5e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGTGCA 24  
|||||  
Db 689 CTAATTAAGACACAAATTGGTGCA 712  
|||||

RESULT 30  
US-10-437-963-21294/c  
Sequence 21294, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 21294  
LENGTH: 1003  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_26579C.1  
US-10-437-963-21294

Query Match 70.4%; Score 17.6; DB 7; Length 1003;  
Best Local Similarity 83.3%; Pred. No. 1.5e+03;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25  
|||||  
Db 956 AAATTAAGTGAATATGCTTGCAA 933  
|||||

RESULT 31  
US-10-739-930-3797/c  
Sequence 3797, Application US/10739930  
Publication No. US20040216190A1  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT  
FILE REFERENCE: 38-21(53377)B  
CURRENT APPLICATION NUMBER: US/10/739,930  
CURRENT FILING DATE: 2003-12-18

```

; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 3797
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: GLYMA-23APR03-CLUSTER73803_6
US-10-739-930-3797

Query Match          70.4%; Score 17.6; DB 8; Length 1431;
Best Local Similarity 83.3%; Pred. No. 1.6e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTTCGTCAA 25
   ||||| ||||| ||||| ||||| |||||
DB 396 AAATTAAAGAACTACTCTTGAAA 373

RESULT 32
US-10-425-114-3484/c
; Sequence 3484, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 3484
; LENGTH: 2726
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700266667_FLI
US-10-425-114-3484

Query Match          70.4%; Score 17.6; DB 7; Length 2726;
Best Local Similarity 83.3%; Pred. No. 1.8e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
DB 528 CAAATCAGAGAGACTTTCGTGCA 505

RESULT 33
US-10-425-115-57374/c
; Sequence 57374, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 57374
; LENGTH: 2828
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_152319C.1
```

```

US-10-425-115-57374

Query Match          70.4%; Score 17.6; DB 8; Length 2828;
Best Local Similarity 83.3%; Pred. No. 1.8e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
DB 528 CAAATCAGAGAGACTTTCGTGCA 505

RESULT 34
US-11-097-143-4939
; Sequence 4939, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4939
; LENGTH: 10892
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-4939

Query Match          70.4%; Score 17.6; DB 10; Length 10892;
Best Local Similarity 83.3%; Pred. No. 2.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAAGAGACTATTTCGTCAA 25
   ||||| ||||| ||||| ||||| |||||
DB 7456 AAATAAATCAGACTATTTCGTCAA 7479

RESULT 35
US-10-282-122A-34396
; Sequence 34396, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
```



```

RESULT 36
US-10-724-972A-326
; Sequence 326, Application US/10724972A
; Publication No. US20040147734A1
; GENERAL INFORMATION:
; APPLICANT: Doucette-Stamm, Lynn
; APPLICANT: Bush, David
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH03-15
; CURRENT APPLICATION NUMBER: US/10/724,972A
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 09/450,969
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 09/134,001
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 326
; LENGTH: 30612
; TYPE: DNA
; ORGANISM: S.epidermidis
US-10-724-972A-326

Query Match          70.4%; Score 17.6; DB 7; Length 30612;
Best Local Similarity 83.3%;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Query Match      70.4%; Score 17.6; DB 10; Length 49380;
Best Local Similarity 83.3%; Pred.No.2.7e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 CAATTTAAAGAGACTATTCTGTGCA 24
          ||| ||||| ||||| |||||
Db      7624 CAATTTAAAGAGCCTGTCTGTGCA 7647

RESULT 38
US-10-085-117-355
; Sequence 355, Application US/10085117
; Publication No. US20030232334A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 355
; LENGTH: 194945
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation

```

```
; LOCATION: (1)...(194945)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-355

Query Match      70.4%; Score 17.6; DB 6; Length 194945;
Best Local Similarity 83.3%; Pred. No. 3.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25
Db 192384 AAATTAAGAGACTATTTCGTGCAA 192407

RESULT 39
US-10-087-192-1282/c
; Sequence 1282, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1282
; LENGTH: 220224
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(220224)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1282

Query Match      70.4%; Score 17.6; DB 5; Length 220224;
Best Local Similarity 83.3%; Pred. No. 3.2e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25
Db 150331 AAATTAAGAGAGGATTGCTTCAA 150308

RESULT 40
US-10-367-094-83
; Sequence 83, Application US/10367094
; Publication No. US20040170982A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc Malandro
; TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
; FILE REFERENCE: 529452001500
; CURRENT APPLICATION NUMBER: US/10/367,094
; CURRENT FILING DATE: 2003-02-14
; NUMBER OF SEQ ID NOS: 203
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83
; LENGTH: 277616
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-367-094-83

Query Match      70.4%; Score 17.6; DB 7; Length 277616;
Best Local Similarity 83.3%; Pred. No. 3.3e+03;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25
Db 160286 AACTTAAGAGATATGTGCAA 160309

RESULT 41
US-10-321-039-122
; Sequence 122, Application US/10321039
; Publication No. US20040014067A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Lukowiak, Andrew
; APPLICANT: Jarvis, Nancy
; APPLICANT: Kurensky, David
; TITLE OF INVENTION: Amplification Methods and Compositions
; FILE REFERENCE: FORS-06960
; CURRENT APPLICATION NUMBER: US/10/321,039
; CURRENT FILING DATE: 2002-12-17
; PRIOR APPLICATION NUMBER: 09/998,157
; PRIOR FILING DATE: 2001-11-30
; PRIOR APPLICATION NUMBER: 60/329,113
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/360,489
; PRIOR FILING DATE: 2001-10-19
; NUMBER OF SEQ ID NOS: 759
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 122
; LENGTH: 1021
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: {551}..{551}
; OTHER INFORMATION: n can be c or t.
US-10-321-039-122

Query Match      69.6%; Score 17.4; DB 7; Length 1021;
Best Local Similarity 94.7%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCG 20
Db 382 AAATTAAGAGACTATTTCG 400

RESULT 42
US-10-052-482-130
; Sequence 130, Application US/10052482
; Publication No. US20040072264A1
; GENERAL INFORMATION:
; APPLICANT: Engelhard, Eric
; APPLICANT: Morris, David
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: A-71087/RMS/DCF
; CURRENT APPLICATION NUMBER: US/10/052,482
; CURRENT FILING DATE: 2002-08-15
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 241
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 130
; LENGTH: 24888
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-052-482-130

Query Match      69.6%; Score 17.4; DB 7; Length 24888;
Best Local Similarity 94.7%; Pred. No. 3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCG 20
```

```

; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-9592
Query Match 68.8%; Score 17.2; DB 3; Length 114;
Best Local Similarity 86.4%; Pred. No. 1.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTGCGT 22
Db 27 CAAATTAAGAGACTATTTCATG 6

RESULT 45
US-10-972-079-58172/c
; Sequence 58172, Application US/10972079
; Publication No. US20050153317A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: ROSENFELD, David
; APPLICANT: KERR, Richard
; APPLICANT: BATES, Stephen
; APPLICANT: HOLM, Tom
; TITLE OF INVENTION: METHODS & SYSTEMS FOR INFERRING TRAITS TO BREED & MANAGE NON-BEE
; TITLE OF INVENTION: LIVESTOCK
; FILE REFERENCE: MM1110-2
; CURRENT APPLICATION NUMBER: US/10/972,079
; CURRENT FILING DATE: 2004-10-22
; PRIOR APPLICATION NUMBER: US 60/514,333
; PRIOR FILING DATE: 2003-10-24
; NUMBER OF SEQ ID NOS: 96631
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58172
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Chicken 19866894314824_1
US-10-972-079-58172

Query Match 68.8%; Score 17.2; DB 9; Length 600;
Best Local Similarity 86.4%; Pred. No. 2.1e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTTCGTGCA 24
Db 525 AATTAAGAGACTATACATTCA 504

RESULT 46
US-10-027-632-114621/c
; Sequence 114621, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
```

```

; OTHER INFORMATION: n equals a,t,g, or c
US-10-034-650-1082
Query Match 69.6%; Score 17.4; DB 6; Length 53106;
Best Local Similarity 94.7%; Pred. No. 3.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCG 20
Db 47351 AAATTAAGAGACTATTTCG 47333

RESULT 44
US-09-783-590-9592/c
; Sequence 9592, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16 2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9592
; LENGTH: 114
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (33)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (39)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (114)
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114621
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114621

Query Match          68.8%; Score 17.2; DB 5; Length 676;
Best Local Similarity 79.2%; Pred. No. 2.2e+03;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
   ||:|||||
Db 357 AAMTTTAAGAGATTATCTAGCAA 334

RESULT 47
US-10-027-632-114622/c
; Sequence 114622, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114622
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114622

Query Match          68.8%; Score 17.2; DB 5; Length 676;
Best Local Similarity 79.2%; Pred. No. 2.2e+03;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
   ||:|||||
Db 357 AAMTTTAAGAGATTATCTAGCAA 334

RESULT 48
US-10-027-632-114623/c
; Sequence 114623, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114623
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114623

Query Match          68.8%; Score 17.2; DB 6; Length 676;
Best Local Similarity 79.2%; Pred. No. 2.2e+03;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
   ||:|||||
Db 357 AAMTTTAAGAGATTATCTAGCAA 334

RESULT 50
US-10-027-632-114622/c
; Sequence 114622, Application US/10027632
; Publication No. US20030204075A9
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114621
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114621

Query Match          68.8%; Score 17.2; DB 5; Length 676;
Best Local Similarity 79.2%; Pred. No. 2.2e+03;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
   ||:|||||
Db 357 AAMTTTAAGAGATTATCTAGCAA 334

RESULT 49
US-10-027-632-114621/c
; Sequence 114621, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 114621
; LENGTH: 676
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-114621

Query Match          68.8%; Score 17.2; DB 6; Length 676;
Best Local Similarity 79.2%; Pred. No. 2.2e+03;
Matches 19; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGCAA 25
   ||:|||||
Db 357 AAMTTTAAGAGATTATCTAGCAA 334

RESULT 50
US-10-027-632-114622/c
; Sequence 114622, Application US/10027632
; Publication No. US20030204075A9
```

Query Match 68.8%; Score 17.2; DB 7; Length 3387;  
Best Local Similarity 86.4%; Pred. No. 2.7e+03;

```
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTCGTG 22
    ||||| ||||| ||||| |||||
Db 1998 CAATTAACGAGACTATTCGTG 1977

RESULT 54
US-10-238-075-507/c
; Sequence 507, Application US/10238075
; Publication No. US20030148324A1
; GENERAL INFORMATION:
; APPLICANT: I.N.S.E.R.M.
; TITLE OF INVENTION: Polynucleotides which are of nature B2/D+ A- and which are isolated
; FILE REFERENCE: BLANDINE
; CURRENT FILING DATE: 2002-09-10
; CURRENT APPLICATION NUMBER: US/10/238,075
; PRIOR APPLICATION NUMBER: 0003145
; PRIOR FILING DATE: 2000-03-10
; NUMBER OF SEQ ID NOS: 1576
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 507
; LENGTH: 12264
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-238-075-507

Query Match 68.8%; Score 17.2; DB 6; Length 12264;
Best Local Similarity 86.4%; Pred. No. 3.3e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGC 23
    ||||| ||||| ||||| |||||
Db 7425 AAATTAATGATCTATTCCTGC 7404

RESULT 55
US-10-085-959-11
; Sequence 11, Application US/10085959
; Publication No. US20030165870A1
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; APPLICANT: Welch, Rodney A.
; APPLICANT: Burland, Valerie D.
; TITLE OF INVENTION: No. US20030165870A1 Sequence of E. Coli CFT073
; FILE REFERENCE: 960296.97648
; CURRENT FILING DATE: 2002-03-01
; CURRENT APPLICATION NUMBER: US/10/085,959
; PRIOR FILING DATE: 2000-10-19
; NUMBER OF SEQ ID NOS: 255
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 23654
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION:
; NAME/KEY: misc feature
; LOCATION: (12189)..(12189)
; OTHER INFORMATION: Unsure
US-10-085-959-11

Query Match 68.8%; Score 17.2; DB 6; Length 23654;
Best Local Similarity 86.4%; Pred. No. 3.6e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCGTGC 23
    ||||| ||||| ||||| |||||
Db 4839 AAATTAATGATCTATTCCTGC 4860
```

```
RESULT 56
US-09-764-877-2308/c
; Sequence 2308, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2308
; LENGTH: 32132
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2308

Query Match 68.8%; Score 17.2; DB 3; Length 32132;
Best Local Similarity 86.4%; Pred. No. 3.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCGTGCA 25
    ||||| ||||| ||||| |||||
Db 26595 ATTAAGAGATAATTCATGCA 26574

RESULT 57
US-09-764-891-9748/c
; Sequence 9748, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9748
; LENGTH: 32132
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9748

Query Match 68.8%; Score 17.2; DB 3; Length 32132;
Best Local Similarity 86.4%; Pred. No. 3.7e+03;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTCGTGCA 25
    ||||| ||||| ||||| |||||
Db 26595 ATTAAGAGATAATTCATGCA 26574

RESULT 58
US-10-242-515-2308/c
; Sequence 2308, Application US/10242515
; Publication No. US20040009488A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005C1
; CURRENT APPLICATION NUMBER: US/10/242,515
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: 09/764,877
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
```

Mon Jan 30 07:42:53 2006

;  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 4031  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2308  
; LENGTH: 32132  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; US-10-242-515-2308

Query Match 68.8%; Score 17.2; DB 6; Length 32132;  
Best Local Similarity 86.4%; Pred. No. 3.7e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTCAA 25  
|||||  
Db 26595 ATTAAGAGATAATTCATGCAA 26574

RESULT 59  
US-10-087-192-1810  
; Sequence 1810, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1810  
; LENGTH: 133632  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-087-192-1810

Query Match 68.8%; Score 17.2; DB 5; Length 133632;  
Best Local Similarity 86.4%; Pred. No. 4.5e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 AATTAAGAGACTATTTCGTGCA 24  
|||||  
Db 22221 AATAAGGAGACTATTTCATGCA 22242

RESULT 60  
US-10-087-192-988/c  
; Sequence 988, Application US/10087192  
; Publication No. US20020182586A1  
; GENERAL INFORMATION:  
; APPLICANT: Morris, David W.  
; APPLICANT: Engelhard, Eric K.

;  
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: CANCER  
; FILE REFERENCE: 529452000122  
; CURRENT APPLICATION NUMBER: US/10/087,192  
; CURRENT FILING DATE: 2002-03-01  
; PRIOR APPLICATION NUMBER: US 09/747,377  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: US 09/798,586  
; PRIOR FILING DATE: 2001-03-02  
; NUMBER OF SEQ ID NOS: 2059  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 988  
; LENGTH: 203264  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-087-192-988

Query Match 68.8%; Score 17.2; DB 5; Length 203264;  
Best Local Similarity 86.4%; Pred. No. 4.6e+03;  
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAGAGACTATTTCGTGCAA 25  
|||||  
Db 98028 ATTAAGAAACCATTAGTGCAA 98007

Search completed: January 28, 2006, 01:45:40  
Job time : 438.708 secs

***This Page Blank (uspto)***



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:15:53 ; Search time 251.966 Seconds  
(without alignments)  
82.419 Million cell updates/sec

Title: US-10-716-005-3  
Perfect score: 25  
Sequence: 1 caaattaaagagactattctgcaa 25

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 6059551 seqs, 415333918 residues

Total number of hits satisfying chosen parameters: 12119102

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 60 summaries

Database : Published Applications NA New.\*  
1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	17.6	70.4	2947	US-10-793-626-3805	Sequence 3805, Ap
2	17.6	70.4	3285	US-10-793-626-3153	Sequence 3153, Ap
3	17.6	70.4	3730	US-10-793-626-3499	Sequence 3499, Ap
C 4	17	68.0	2131	US-10-750-185-30811	Sequence 30811, A
C 5	17	68.0	2131	US-10-750-623-30811	Sequence 30811, A
C 6	16.6	66.4	2076	US-10-632-150-9	Sequence 9, Appli
C 7	16.6	66.4	2076	US-11-073-457-9	Sequence 9, Appli
C 8	16.6	66.4	2076	US-11-073-450-9	Sequence 9, Appli
C 9	16.6	66.4	50353	US-10-995-561-13305	Sequence 13305, A
C 10	16.6	66.4	129021	US-11-117-187-202	Sequence 202, App
11	16.6	66.4	145857	US-11-121-086-34	Sequence 34, Appli
12	16.4	65.6	201	US-10-995-561-58262	Sequence 58262, A
C 13	16.4	65.6	201	US-10-995-561-80898	Sequence 80898, A
C 14	16.4	65.6	1719	US-10-750-185-48305	Sequence 48305, A
C 15	16.4	65.6	1719	US-10-750-623-48305	Sequence 48305, A
C 16	16.4	65.6	2941	US-10-750-185-53117	Sequence 53117, A
17	16.4	65.6	2941	US-10-750-623-53117	Sequence 53117, A
C 18	16.4	65.6	21852	US-10-995-561-13492	Sequence 13492, A
C 19	16.4	65.6	43985	US-10-995-561-13337	Sequence 13337, A
C 20	16.2	64.8	201	US-10-995-561-15984	Sequence 15984, A
C 21	16.2	64.8	201	US-10-995-561-16055	Sequence 16055, A
C 22	16.2	64.8	429	US-11-136-527-1338	Sequence 1338, Ap

RESULT 1

US-10-793-626-3805  
; Sequence 3805, Application US/10793626  
; Publication No. US20050255478A1

; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN

; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258

; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ IDS NOS: 4472

; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3805

; LENGTH: 2947  
; TYPE: DNA

; ORGANISM: Artificial Sequence  
; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: nucleic acid sequence

US-10-793-626-3805

Query Match 70.4%; Score 17.6; DB 7; Length 2947;  
Best Local Similarity 83.3%; Pred. No. 54;  
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTCTGCA 24

Db 1894 CAAATTAAGAGACTATTCTGCA 1917

Sequence 5434, Ap  
Sequence 6155, Ap  
Sequence 2059, Ap  
Sequence 1, Appli  
Sequence 13201, A  
Sequence 59, Appli  
Sequence 58, Appli  
Sequence 13233, A  
Sequence 13233, A  
Sequence 56986, A  
Sequence 56986, A  
Sequence 31221, A  
Sequence 31221, A  
Sequence 39461, A  
Sequence 39461, A  
Sequence 41491, A  
Sequence 41491, A  
Sequence 35978, A  
Sequence 35978, A  
Sequence 54406, A  
Sequence 54406, A  
Sequence 51187, A  
Sequence 51187, A  
Sequence 82, Appli  
Sequence 663, Appli  
Sequence 20, Appli  
Sequence 20, Appli  
Sequence 189, Appli  
Sequence 189, Appli  
Sequence 2881, Ap  
Sequence 61, Appli  
Sequence 29, Appli  
Sequence 60, Appli  
Sequence 54, Appli  
Sequence 13495, A  
Sequence 95, Appli  
Sequence 1, Appli  
Sequence 181, Appli

ALIGNMENTS

```
RESULT 2
US-10-793-626-3153
; Sequence 3153, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3153
; LENGTH: 3285
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3153
Query Match 70.4%; Score 17.6; DB 7; Length 3285;
Best Local Similarity 83.3%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
||| ||||| ||| |||||
Db 1447 CAAATAAAGAGAAATAAGTGCA 1470

RESULT 3
US-10-793-626-3499
; Sequence 3499, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3499
; LENGTH: 3730
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-10-793-626-3499
Query Match 70.4%; Score 17.6; DB 7; Length 3730;
Best Local Similarity 83.3%; Pred. No. 55;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24
||| ||||| ||| |||||
Db 1892 CAAATAAAGAGAAATAAGTGCA 1915

RESULT 4
US-10-750-185-30811/c
; Sequence 30811, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30811
; LENGTH: 2131
; TYPE: DNA
; ORGANISM: Bovine
; ORGANISM: Bovine 19866880964537
US-10-750-185-30811
Query Match 68.0%; Score 17; DB 7; Length 2131;
Best Local Similarity 80.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 25
||| ||||| ||| |||||
Db 40 CAAATTAAGAGACTATTTCGTGCTA 16

RESULT 5
US-10-750-623-30811/c
; Sequence 30811, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30811
; LENGTH: 2131
; TYPE: DNA
; ORGANISM: Bovine
; ORGANISM: Bovine 19866880964537
US-10-750-623-30811
Query Match 68.0%; Score 17; DB 7; Length 2131;
Best Local Similarity 80.0%; Pred. No. 98;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 25
||| ||||| ||| |||||
Db 40 CAAATTAAGAGACTATTTCGTGCTA 16

RESULT 6
US-10-632-150-9/c
; Sequence 9, Application US/10632150
; Publication No. US20050251871A1
; GENERAL INFORMATION:
; APPLICANT: Chiau, D.
; APPLICANT: Pagano, M.
; APPLICANT: Latres, E.
; TITLE OF INVENTION: NOVEL UBIQUITIN LIGASES AS THERAPEUTIC TARGETS
; FILE REFERENCE: 5914-081
```

; CURRENT APPLICATION NUMBER: US/10/632,150  
; CURRENT FILING DATE: 2003-07-30  
; PRIOR APPLICATION NUMBER: US/09/385,219  
; PRIOR FILING DATE: 1999-08-27  
; PRIOR APPLICATION NUMBER: 60/098,355  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: 60/118,568  
; PRIOR FILING DATE: 1999-02-03  
; PRIOR APPLICATION NUMBER: 60/124,449  
; PRIOR FILING DATE: 1999-03-15  
; NUMBER OF SEQ ID NOS: 90  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 2076  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-632-150-9

Query Match 66.4%; Score 16.6; DB 7; Length 2076;  
Best Local Similarity 82.6%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
DB 1298 AATTACAGAGACACTTCGTACAA 1276

RESULT 7  
US-11-073-457-9/c  
; Sequence 9, Application US/11073457  
; Publication No. US20050260556A1  
; GENERAL INFORMATION:  
; APPLICANT: Pagano, M.  
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF PROLIFERATIVE DISORDERS  
; FILE REFERENCE: 5914-090-999  
; CURRENT APPLICATION NUMBER: US/11/073,457  
; CURRENT FILING DATE: 2005-03-04  
; PRIOR APPLICATION NUMBER: 10/042,417  
; PRIOR FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 60/260,179  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 2076  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-073-457-9

Query Match 66.4%; Score 16.6; DB 8; Length 2076;  
Best Local Similarity 82.6%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
DB 1298 AATTACAGAGACACTTCGTACAA 1276

RESULT 8  
US-11-073-460-9/c  
; Sequence 9, Application US/11073460  
; Publication No. US20050272066A1  
; GENERAL INFORMATION:  
; APPLICANT: Pagano, M.  
; TITLE OF INVENTION: METHODS TO IDENTIFY COMPOUNDS USEFUL FOR THE TREATMENT OF PROLIFERATIVE DISORDERS  
; FILE REFERENCE: 5914-090-999  
; CURRENT APPLICATION NUMBER: US/11/073,460  
; CURRENT FILING DATE: 2005-03-04  
; PRIOR APPLICATION NUMBER: 10/042,417  
; PRIOR FILING DATE: 2002-01-07  
; PRIOR APPLICATION NUMBER: 60/260,179

; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 2076  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-073-460-9

Query Match 66.4%; Score 16.6; DB 8; Length 2076;  
Best Local Similarity 82.6%; Pred. No. 1.5e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
DB 1298 AATTACAGAGACACTTCGTACAA 1276

RESULT 9  
US-10-995-561-13305/c  
; Sequence 13305, Application US/10995561  
; Publication No. US20050272054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13305  
; LENGTH: 50353  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-13305

Query Match 66.4%; Score 16.6; DB 7; Length 50353;  
Best Local Similarity 82.6%; Pred. No. 2.3e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 AATTAAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
DB 16515 AATAAAAGAGATTATCGGTGTA 16493

RESULT 10  
US-11-117-187-202  
; Sequence 202, Application US/11117187  
; Publication No. US20050266560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: ARCD:309US  
; CURRENT APPLICATION NUMBER: US/11/117,187  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: US/09/531,120  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,219  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 202  
; LENGTH: 129021  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-11-117-187-202

Query Match 66.4%; Score 16.6; DB 8; Length 129021;  
Best Local Similarity 82.6%; Pred. No. 2.6e+02;  
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 2 AAATTAAGAGACTATTCTGCA 24
||||| ||| ||| ||| |||
Db 80808 AAATTAAGAGAAATTTCTGCA 80830

RESULT 11
US-11-121-086-34
; Sequence 34, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: NIELSEN, KIRSTEN V.
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.8000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 34
; LENGTH: 165857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-34

Query Match 66.4%; Score 16.6; DB 8; Length 165857;
Best Local Similarity 82.6%; Pred. No. 2.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTCTGCA 24
||||| ||| ||| ||| |||
Db 61880 AAATTAAGGAGCTGTCAAGCA 61902

RESULT 12
US-10-995-561-58262
; Sequence 58262, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58262
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-58262

Query Match 65.6%; Score 16.4; DB 7; Length 201;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATT 18
||||| ||| ||| ||| |||
Db 8 CAAATTAAGAGACTACT 25

RESULT 13
US-10-995-561-80898/c
; Sequence 80898, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
```

```
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80898
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-80898
```

```
Query Match 65.6%; Score 16.4; DB 7; Length 201;
Best Local Similarity 94.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATT 18
||||| ||| ||| ||| |||
Db 194 CAAATTAAGAGACTACT 177
```

```
RESULT 14
US-10-750-185-48305/c
; Sequence 48305, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48305
; LENGTH: 1719
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-48305
```

```
Query Match 65.6%; Score 16.4; DB 7; Length 1719;
Best Local Similarity 94.4%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTC 19
||||| ||| ||| ||| |||
Db 289 AAATTAAGGAGACTATTTC 272
```

```
RESULT 15
US-10-750-623-48305/c
; Sequence 48305, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
```

; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 48305  
; LENGTH: 1719  
; TYPE: DNA  
; ORGANISM: Bovine 19866880887302  
US-10-750-623-48305

Query Match 65.6%; Score 16.4; DB 7; Length 1719;  
Best Local Similarity 94.4%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTC 19  
Db 289 AAATTAAGAGACTATTC 272

RESULT 16  
US-10-750-185-53117  
; Sequence 53117, Application US/10750185  
; Publication No. US20050260603A1

; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 53117  
; LENGTH: 2941  
; TYPE: DNA  
; ORGANISM: Bovine 19866880648156  
US-10-750-185-53117

Query Match 65.6%; Score 16.4; DB 7; Length 2941;  
Best Local Similarity 94.4%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTC 19  
Db 94 AAATTAAGAGACTATTC 111

RESULT 17  
US-10-750-623-53117  
; Sequence 53117, Application US/10750623  
; Publication No. US20050287531A1

; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1

; SEQ ID NO 53117  
; LENGTH: 2941  
; TYPE: DNA  
; ORGANISM: Bovine 19866880648156  
US-10-750-623-53117

Query Match 65.6%; Score 16.4; DB 7; Length 2941;  
Best Local Similarity 94.4%; Pred. No. 1.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTC 19  
Db 94 AAATTAAGAGACTATTC 111

RESULT 18  
US-10-995-561-13492/c  
; Sequence 13492, Application US/10995561  
; Publication No. US20050272054A1

; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 13492  
; LENGTH: 21852  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-13492

Query Match 65.6%; Score 16.4; DB 7; Length 21852;  
Best Local Similarity 94.4%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATT 18  
Db 10295 CAAATTAAGAGACTACT 10278

RESULT 19  
US-10-995-561-13337  
; Sequence 13337, Application US/10995561  
; Publication No. US20050272054A1

; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 13337  
; LENGTH: 43985  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-995-561-13337

Query Match 65.6%; Score 16.4; DB 7; Length 43985;  
Best Local Similarity 94.4%; Pred. No. 2.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATT 18  
Db 1616 CAAATTAAGAGACTACT 1633

```
RESULT 20
US-10-995-561-15984/c
; Sequence 15984, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15984
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-15984
Query Match 64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 TTAAGAGACTATTTCGTCAA 25
||| ||||| ||||| |||||
Db 199 TTTAGAGACTATTTCGTCAA 179

RESULT 21
US-10-995-561-16055/c
; Sequence 16055, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16055
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-16055
Query Match 64.8%; Score 16.2; DB 7; Length 201;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 TTAAGAGACTATTTCGTCAA 25
||| ||||| ||||| |||||
Db 198 TTTAGAGACTATTTCGTCAA 178

RESULT 22
US-11-136-527-1338/c
; Sequence 1338, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1338
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-1338
Query Match 64.8%; Score 16.2; DB 8; Length 429;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTG 22
||||| ||||| ||||| |||||
Db 40 AAATTAAGAGCTCTATGCATG 20

RESULT 23
US-11-136-527-5434
; Sequence 5434, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5434
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-5434
Query Match 64.8%; Score 16.2; DB 8; Length 429;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTG 22
||||| ||||| ||||| |||||
Db 390 AAATTAAGAGCTCTATGCATG 410

RESULT 24
US-11-136-527-6155
; Sequence 6155, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6155
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6155
Query Match 64.8%; Score 16.2; DB 8; Length 1400;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGT 21
||||| ||||| ||||| |||||
Db 1282 CAAATTAAGAGCTATTTCCT 1302
```

```

RESULT 25
US-11-136-527-2059
; Sequence 2059, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2059
; LENGTH: 1656
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; US-11-136-527-2059

Query Match          64.8%; Score 16.2; DB 8; Length 1656;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGT 21
||||| ||||| ||||| ||||| |||||
DB 1538 CAAATAAATAGACTATTCTT 1558

RESULT 26
US-11-121-634-1/c
; Sequence 1, Application US/11121634
; Publication No. US20050282198A1
; GENERAL INFORMATION:
; APPLICANT: Duff, Gordon
; TITLE OF INVENTION: Diagnostics and Therapeutics for Diseases Associated with an IL-1
; FILE REFERENCE: 24299-508CON3 CIP1
; CURRENT APPLICATION NUMBER: US/11/121,634
; CURRENT FILING DATE: 2005-05-03
; PRIOR APPLICATION NUMBER: US 10/802,061
; PRIOR FILING DATE: 2004-03-15
; PRIOR APPLICATION NUMBER: US 09/845,129
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 09/345,217
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/GB98/01481
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: GB9711040.7
; PRIOR FILING DATE: 1997-05-29
; PRIOR APPLICATION NUMBER: US 10/300,011
; PRIOR FILING DATE: 2002-11-19
; PRIOR APPLICATION NUMBER: US 60/386,020
; PRIOR FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: US 60/331,681
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: US 60/567,727
; PRIOR FILING DATE: 2004-05-03
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 11970
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-121-634-1

Query Match          64.8%; Score 16.2; DB 8; Length 11970;
Best Local Similarity 85.7%; Pred. No. 2.9e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGT 21
||||| ||||| ||||| ||||| |||||
DB 88036 CAAATAAAGTACTATTTCAT 88016

RESULT 27
US-10-995-561-13201/c
; Sequence 13201, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13201
; LENGTH: 23107
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-995-561-13201

Query Match          64.8%; Score 16.2; DB 7; Length 23107;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 TTAAGAGACTATTTCGTGCAA 25
||||| ||||| ||||| ||||| |||||
DB 11488 TTTTAGAGACTATTTGTGCAA 11468

RESULT 28
US-11-112-908-59/c
; Sequence 59, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 59
; LENGTH: 168656
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-11-112-908-59

Query Match          64.8%; Score 16.2; DB 8; Length 168656;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTTCGT 21
||||| ||||| ||||| ||||| |||||
DB 88036 CAAATAAAGTACTATTTCAT 88016

RESULT 29
US-11-112-908-58/c
; Sequence 58, Application US/11112908
```

```
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 58
; LENGTH: 170285
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-58

Query Match      64.8%; Score 16.2; DB 8; Length 170285;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGT 21
Db 88041 CAAATTAAGTCACTATTTCAT 88021

RESULT 30
US-10-995-561-13233
; Sequence 13233, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13233
; LENGTH: 199130
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)_(199130)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13233

Query Match      64.8%; Score 16.2; DB 7; Length 199130;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAAGAGACTATTTCGTGCA 24
Db 82228 ATTGAAGAGACTATTCTTTCA 82248

RESULT 31
US-10-995-561-13233/c
; Sequence 13233, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13233
; LENGTH: 199130
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)_(199130)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13233

Query Match      64.8%; Score 16.2; DB 7; Length 199130;
Best Local Similarity 85.7%; Pred. No. 4.1e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ATTAAAGAGACTATTTCGTGCA 24
Db 116275 ATTGAAGAGACTATTCTTTCA 116255

RESULT 32
US-10-750-185-56986
; Sequence 56986, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56986
; LENGTH: 940
; TYPE: DNA
; ORGANISM: Bovine
US-10-750-185-56986

Query Match      64.0%; Score 16; DB 7; Length 940;
Best Local Similarity 79.2%; Pred. No. 2.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAATTAAGAGACTATTTCGTGCA 24
Db 578 CATATTAAAAAGATTATTGTGAA 601

RESULT 33
US-10-750-623-56986
; Sequence 56986, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
```





```
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 39461
; LENGTH: 1646
; TYPE: DNA
; ORGANISM: Bovine 19866880785007
US-10-750-623-39461

Query Match          64.0%; Score 16; DB 7; Length 1646;
Best Local Similarity 79.2%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 1322 CAAATTAAAGAGATTATGGGAGAA 1299

RESULT 38
US-10-750-185-41491/c
; Sequence 41491, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41491
; LENGTH: 1699
; TYPE: DNA
; ORGANISM: Bovine 19866881235275
US-10-750-185-41491

Query Match          64.0%; Score 16; DB 7; Length 1699;
Best Local Similarity 79.2%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 1322 CAAATTAAAGAGATTATGGGAGAA 1299

RESULT 39
US-10-750-623-41491/c
; Sequence 41491, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41491
; LENGTH: 1699
; TYPE: DNA
; ORGANISM: Bovine 19866880908844
US-10-750-185-35978/c
; Sequence 35978, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 35978
; LENGTH: 2060
; TYPE: DNA
; ORGANISM: Bovine 19866880908844
US-10-750-185-35978

Query Match          64.0%; Score 16; DB 7; Length 2060;
Best Local Similarity 79.2%; Pred. No. 2.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAAGAGACTATTCGTGCA 24
   ||||| ||||| ||||| ||||| |||||
Db 1033 CAAATTACAGACCATTGTGCA 1010

RESULT 41
US-10-750-623-35978/c
; Sequence 35978, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 35978
; LENGTH: 2060
; TYPE: DNA
; ORGANISM: Bovine 19866880908844
```

US-10-750-623-35978

Query Match 64.0%; Score 16; DB 7; Length 2060;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CAAATTAAGAGACTATTTCGTGCA 24  
||||| ||||| ||||| ||||| |||||  
Db 1033 CAAATTACAGAGACCAATTGGTCA 1010

RESULT 42

US-10-750-185-54406/c  
; Sequence 54406, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 54406  
; LENGTH: 2342  
; TYPE: DNA  
; ORGANISM: Bovine 19866880508883  
US-10-750-185-54406

Query Match 64.0%; Score 16; DB 7; Length 2342;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 1799 AATTTAAGAGCTATTTGAGCAA 1776

RESULT 43

US-10-750-623-54406/c  
; Sequence 54406, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 54406  
; LENGTH: 2342  
; TYPE: DNA  
; ORGANISM: Bovine 19866880508883  
US-10-750-623-54406

Query Match 64.0%; Score 16; DB 7; Length 2342;

Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 1799 AATTTAAGAGCTATTTGAGCAA 1776

RESULT 44

US-10-750-185-51187  
; Sequence 51187, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 51187  
; LENGTH: 2509  
; TYPE: DNA  
; ORGANISM: Bovine 19866881027540  
US-10-750-185-51187

Query Match 64.0%; Score 16; DB 7; Length 2509;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AAATTAAGAGACTATTTCGTGCAA 25  
||||| ||||| ||||| ||||| |||||  
Db 2455 AAATCAACAGACTATTCATTGAA 2478

RESULT 45

US-10-750-623-51187  
; Sequence 51187, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 51187  
; LENGTH: 2509  
; TYPE: DNA  
; ORGANISM: Bovine 19866881027540  
US-10-750-623-51187

Query Match 64.0%; Score 16; DB 7; Length 2509;  
Best Local Similarity 79.2%; Pred. No. 2.9e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;



Query Match 64.0%; Score 16; DB 8; Length 48000;  
Best Local Similarity 79.2%; Pred. No. 4.4e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCA 25  
Db 15508 ATATGAAGAGCGCTATTACTGCA 15485

RESULT 50  
US-11-117-187-189/c  
; Sequence 189, Application US/11117187  
; Publication No. US2005026560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; APPLICANT: COPENHAVER, GREGORY  
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS  
; FILE REFERENCE: ARCD:309US  
; CURRENT APPLICATION NUMBER: US/11/117,187  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: US/09/531,120  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,219  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 189  
; LENGTH: 83391  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: modified base  
; LOCATION: (2556)..(2605)  
; OTHER INFORMATION: N = A, C G, or T/U  
US-11-117-187-189

Query Match 64.0%; Score 16; DB 8; Length 83391;  
Best Local Similarity 79.2%; Pred. No. 4.6e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTTCGTGCA 24  
Db 5781 CAACTTGAAGAGCGATTTCGGCA 5758

RESULT 51  
US-11-117-187-195/c  
; Sequence 195, Application US/11117187  
; Publication No. US2005026560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; APPLICANT: COPENHAVER, GREGORY  
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS  
; FILE REFERENCE: ARCD:309US  
; CURRENT APPLICATION NUMBER: US/11/117,187  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: US/09/531,120  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,219  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 195  
; LENGTH: 90336  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-11-117-187-195

Query Match 64.0%; Score 16; DB 8; Length 90336;  
Best Local Similarity 79.2%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAGAGACTATTTCGTGCA 24

Db 18537 CAACTTGAAGAGCGATTTCGGCA 18514

RESULT 52  
US-11-124-368A-2881/c  
; Sequence 2881, Application US/11124368A  
; Publication No. US20050287559A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: James J. Devlin  
; APPLICANT: May Luke  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001524  
; CURRENT APPLICATION NUMBER: US/11/124,368A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,845  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/625,936  
; PRIOR FILING DATE: 2004-11-09  
; NUMBER OF SEQ ID NOS: 21112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2881  
; LENGTH: 100000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-124-368A-2881

Query Match 64.0%; Score 16; DB 8; Length 100000;  
Best Local Similarity 79.2%; Pred. No. 4.7e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCA 25  
Db 54468 AAATTAAGAACTATTGATGAA 54445

RESULT 53  
US-11-112-908-61/c  
; Sequence 61, Application US/11112908  
; Publication No. US20050260659A1  
; GENERAL INFORMATION:  
; APPLICANT: Harris, Cole  
; APPLICANT: Davis, Lisa M.  
; TITLE OF INVENTION: Breast Cancer Biomarkers  
; FILE REFERENCE: 04-164-US  
; CURRENT APPLICATION NUMBER: US/11/112,908  
; CURRENT FILING DATE: 2005-04-22  
; PRIOR APPLICATION NUMBER: US 60/564,758  
; PRIOR FILING DATE: 2004-04-23  
; PRIOR APPLICATION NUMBER: US 60/575,978  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/631,702  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/633,826  
; PRIOR FILING DATE: 2004-12-07  
; NUMBER OF SEQ ID NOS: 511  
; SOFTWARE: Patentin version 3.3  
; SEQ ID NO 61  
; LENGTH: 159497  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-112-908-61

Query Match 64.0%; Score 16; DB 8; Length 159497;  
Best Local Similarity 79.2%; Pred. No. 5e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTTCGTGCA 25  
Db 155485 AAATAAAGAGATTATTCTTCAA 155462

```
RESULT 54
US-11-121-086-29
; Sequence 29, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 29
; LENGTH: 160226
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-29

Query Match 64.0%; Score 16; DB 8; Length 160226;
Best Local Similarity 79.2%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAAGAGACTATTCTGCGAA 25
Db 83051 AAAAAAAGAGACTATTTTGCTA 83074

RESULT 55
US-11-112-908-60/c
; Sequence 60, Application US/11112908
; Publication No. US20050260659A1
; GENERAL INFORMATION:
; APPLICANT: Harris, Cole
; TITLE OF INVENTION: Breast Cancer Biomarkers
; FILE REFERENCE: 04-164-US
; CURRENT APPLICATION NUMBER: US/11/112,908
; CURRENT FILING DATE: 2005-04-22
; PRIOR APPLICATION NUMBER: US 60/564,758
; PRIOR FILING DATE: 2004-04-23
; PRIOR APPLICATION NUMBER: US 60/575,978
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/631,702
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: US 60/633,826
; PRIOR FILING DATE: 2004-12-07
; NUMBER OF SEQ ID NOS: 511
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 60
; LENGTH: 171427
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-112-908-60

Query Match 64.0%; Score 16; DB 8; Length 171427;
Best Local Similarity 79.2%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAAGAGACTATTCTGCGAA 25
Db 61621 AAATAAAGAGACTATTTCTTCA 61598

RESULT 56
US-11-121-086-54/c
; Sequence 54, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
```

```
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 54
; LENGTH: 189252
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-54

Query Match 64.0%; Score 16; DB 8; Length 189252;
Best Local Similarity 79.2%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAAGAGACTATTCTGCGAA 25
Db 31645 AAAGTAAAGGAGATTATTCATGGA 31622

RESULT 57
US-10-995-561-13495/c
; Sequence 13495, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13495
; LENGTH: 195235
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13495

Query Match 64.0%; Score 16; DB 7; Length 195235;
Best Local Similarity 79.2%; Pred. No. 5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 CAAATTAAAGAGACTATTCTGCGCA 24
Db 173960 CAAAGTAAAGAGACAATCTGTGAA 173937

RESULT 58
US-11-121-086-95/c
; Sequence 95, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 95
; LENGTH: 212716
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-95
```



```

; NAME/KEY: allele
; LOCATION: (1047134)
; OTHER INFORMATION: the 'n' at position 1047134 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1047159)
; OTHER INFORMATION: the 'n' at position 1047159 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1047378)
; OTHER INFORMATION: the 'n' at position 1047378 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1047739)
; OTHER INFORMATION: the 'n' at position 1047739 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1050133)..(1050137)
; OTHER INFORMATION: "ttaa" may be deleted at this position
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1050539)
; OTHER INFORMATION: the 'n' at position 1050539 may be 'c' or 't'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1062808)
; OTHER INFORMATION: the 'n' at position 1062808 may be 'c' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1066392)
; OTHER INFORMATION: the 'n' at position 1066392 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: allele
; LOCATION: (1073711)
; OTHER INFORMATION: the 'n' at position 1073711 may be 'c' or 't'
;
US-10-928-446A-1

```

Query Match 64.0%; Score 16; DB 7; Length 1080000;  
 Best Local Similarity 79.2%; Pred. No. 4.5e+02;  
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGCAA 25  
 Db 623021 AAAATAAGGAATATTGTGCAA 623044

```

RESULT 60
US-10-928-446A-181
; Sequence 181, Application US/10928446A
; Publication No. US20050277123A1
; GENERAL INFORMATION:
; APPLICANT: UNIVERSITY OF UTAH RESEARCH FOUNDATION
; TITLE OF INVENTION: VARIANTS OF NEDD4L ASSOCIATED WITH HYPERTENSION AND
; FILE REFERENCE: 0274-5785.1US
; CURRENT APPLICATION NUMBER: US/10/928,446A
; CURRENT FILING DATE: 2004-08-26
; PRIOR APPLICATION NUMBER: 60/359,741
; PRIOR FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 202
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 181
; LENGTH: 1080000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (826985)..(827008)
; OTHER INFORMATION: full exon 1 range is 826667-827008
; FEATURE:
; NAME/KEY: allele
; LOCATION: (827008)..(827008)

```

```

; OTHER INFORMATION: the 'n' at position 827008 may be 'a' or 'g'
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (843242)..(843315)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (922549)..(922630)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (926021)..(926059)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (929123)..(929176)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (993104)..(993154)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (999547)..(999608)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1000354)..(1000456)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1002118)..(1002284)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1006117)..(1006249)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1007860)..(1008036)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1010940)..(1011014)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1018160)..(1018291)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1018800)..(1018919)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1020028)..(1020225)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1026659)..(1026736)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1028113)..(1028167)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1034316)..(1034374)
; OTHER INFORMATION: exon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1041390)..(1041455)
; OTHER INFORMATION: exon

```



Mon Jan 30 07:42:53 2006

```
FEATURE:
NAME/KEY: CDS
LOCATION: (1043121)..(1043350)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1044868)..(1044989)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1047519)..(1047589)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1050296)..(1050391)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1060368)..(1060441)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1062648)..(1062708)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1064561)..(1064620)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1066207)..(1066314)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1067768)..(1067864)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1068609)..(1068681)
OTHER INFORMATION: exon
FEATURE:
NAME/KEY: CDS
LOCATION: (1073289)..(1073388)
OTHER INFORMATION: full exon 30 range is 1073289-1075279
US-10-928-446A-181
```

Query Match 64.0%; Score 16; DB 7; Length 1080000;  
Best Local Similarity 79.2%; Pred. No. 4.5e+02;  
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AAATTAAGAGACTATTCGTGCAA 25  
Db 623021 AAAATAAGGAGATATTGATGCAA 623044

Search completed: January 28, 2006, 02:00:47  
Job time : 259.216 secs

**This Page Blank (uspto)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:49:12 ; Search time 1620.9 Seconds

(without alignments)  
577.298 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaaatgcagaaacagg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_gse1:\*  
10: gb\_gse2:\*  
11: gb\_gse3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18.4	92.0	163	5	BU581563
C 2	18.4	92.0	482	2	BF654954
C 3	18.4	92.0	582	10	CZ263299
4	18.4	92.0	591	3	BP159851
5	18.4	92.0	599	3	BI770737
6	18	90.0	484	9	AQ312469
7	17.4	87.0	183	7	CK593526
8	17.4	87.0	197	9	AZ464666
9	17.4	87.0	274	1	AL649495
C 10	17.4	87.0	336	1	AL679022
11	17.4	87.0	350	5	EX689357
12	17.4	87.0	412	2	BE666884
C 13	17.4	87.0	429	1	AW035216
C 14	17.4	87.0	429	9	BB9397
C 15	17.4	87.0	433	1	AL966115
C 16	17.4	87.0	438	9	AQ636375
C 17	17.4	87.0	472	3	BP909630
C 18	17.4	87.0	473	5	BW554910
C 19	17.4	87.0	478	2	BF409425
C 20	17.4	87.0	478	3	BP906911
C 21	17.4	87.0	480	9	AZ961721
C 22	17.4	87.0	489	1	AL782393

C 23	17.4	87.0	490	7	CO335661
C 24	17.4	87.0	491	9	AQ013410
C 25	17.4	87.0	505	1	AL672590
C 26	17.4	87.0	525	3	BP908357
C 27	17.4	87.0	526	1	AL655162
28	17.4	87.0	541	10	CZ291235
29	17.4	87.0	550	3	BP131868
30	17.4	87.0	561	3	BP721011
C 31	17.4	87.0	561	9	BZ292732
C 32	17.4	87.0	562	9	AZ596869
C 33	17.4	87.0	564	10	AG292825
34	17.4	87.0	568	10	CS03452
35	17.4	87.0	571	8	CX397388
C 36	17.4	87.0	572	8	CK721858
C 37	17.4	87.0	576	7	CK721858
C 38	17.4	87.0	586	1	AL804540
C 39	17.4	87.0	586	7	CN109679
C 40	17.4	87.0	600	5	BU919936
C 41	17.4	87.0	604	1	AL644275
C 42	17.4	87.0	604	2	BG634601
C 43	17.4	87.0	607	10	CZ498169
C 44	17.4	87.0	610	5	BX764191
C 45	17.4	87.0	613	1	AL793267
C 46	17.4	87.0	618	1	AL967548
C 47	17.4	87.0	619	3	BP685102
C 48	17.4	87.0	620	1	AL661458
C 49	17.4	87.0	620	1	AL783961
C 50	17.4	87.0	621	5	BQ521386
C 51	17.4	87.0	623	10	AG931878
C 52	17.4	87.0	634	8	CX466293
C 53	17.4	87.0	638	7	CN109680
54	17.4	87.0	638	9	AZ600504
55	17.4	87.0	639	5	BX758905
C 56	17.4	87.0	649	1	AL895824
57	17.4	87.0	652	7	CR426638
C 58	17.4	87.0	658	7	CN084914
59	17.4	87.0	663	3	BJ076021
60	17.4	87.0	669	8	CX491676

RESULT 1

BU581563

LOCUS

DEFINITION

BU581563 163 bp mRNA linear EST 17-SEP-2002  
in38c09.y1 Human Fetal Pancreas 1B Homo sapiens cDNA 5' similar to  
SW:MITC2 HUMAN P14209 T-CELL SURFACE GLYCOPROTEIN E2 PRECURSOR ;,  
mRNA sequence.

ACCESSION

BU581563

VERSION

BU581563.1 GI:23066823

KEYWORDS

EST.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 163)

REFERENCE

AUTHORS

Unpublished (2000)

TITLE

JOURNAL

COMMENT

Other ESTs: in38c09.x1

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

ALIGNMENTS

```

Fax: 617-495-8557
Email: dmlton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@imgate.wustl.edu)
Trace considered overall poor quality
Seq primer: -40RP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1..163
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="Fetal Pancreas (4 Pooled Donors, 18 - 20
weeks, Stratagene #738023)"
/dev_stage="Fetal Pancreas"
/clone_lib="Human Fetal Pancreas 18"
/note="Vector: pBluescript SK(-); Site 1: NotI; Site 2:
XhoI; cDNA made by oligo-dT priming. Size-selected on
agarose gel. Average insert size ~1kb. 5' XhoI site was
destroyed after directional cloning. Amplified once.
Contact information: Hiroshi Inoue, MD, Metabolism Div.
(Alan Permutt Lab), Washington University School of
Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO
63110 E-mail: hinoue@imgate.wustl.edu Tel: 314-362-1916,
Fax:314-747-2692."

FEATURES
source
Query Match 92.0%; Score 18.4; DB 5; Length 163;
Best Local Similarity 95.0%; Pred. No. 7.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCGAAGAACAGG 20
|||||
Db 113 CAAGAAATGCGAAGAACAGG 132

RESULT 2
BF654954/c
LOCUS 279320 MARC 3BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
DEFINITION BF654954
ACCESSION BF654954
VERSION BF654954.1 GI:11920086
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 482)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-McKown,C.G., Ferrea,G., Holt,I., Karanycheva,S., Liang,F.,
Quackenbush,J. and Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
11282978
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACACTATGACCAT
BACKWARD: GTTTCACAGTCACGACG
Plate: 74 row: E column: 11

Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..482
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 3BOV"
/note="Vector: pCMV SPORT6; Site1: NotI; Site 2: SalI;
Library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

ORIGIN
Query Match 92.0%; Score 18.4; DB 2; Length 482;
Best Local Similarity 95.0%; Pred. No. 8.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAATGCGAAGAACAGG 20
|||||
Db 289 CAAGCAATGCGAAGAACAGG 270

RESULT 3
CZ263299/c
LOCUS CZ263299
DEFINITION OA_BBA0107F18.f OA_BBA Oryza alta genomic clone OA_BBA0107F18 5',
genomic survey sequence.
ACCESSION CZ263299
VERSION CZ263299.1 GI:60332075
KEYWORDS GSS.
SOURCE Oryza alta
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 582)
SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wisotski,M., Yost,D.,
Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
OMAP (Oryza Map Alignment Project) - Purdue University
Unpublished (2004)
Contact: Scott A. Jackson
Jackson Laboratory
Purdue University
915 W. State St., West Lafayette, IN 47907, USA
Tel: 7654963621
Fax: 7654967255
Email: sjackson@purdue.edu
Basecalling by phred version 0.020425.c. This sequence was derived
from the raw sequence read by clipping with lcy version 1.198.
Bases 36-617 of the raw sequence (length 1085) were retained after
clipping.
Plate: 0107 row: F column: 18
Seq primer: TAA TAC GAC TCA CTA TAG GG
Class: BAC ends.
Location/Qualifiers
1..582
/organism="Oryza alta"
/mol_type="genomic DNA"
/db_xref="taxon:52545"
/clone="OA_BBA0107F18"
/tissue_type="young leaves"
/lab_host="DH10B-T1 phage resistant"
/clone_lib="OA_BBA"
/note="Vector: pAGIBAC1; Site_1: HindIII; Site_2: HindIII"

FEATURES
source
Query Match 92.0%; Score 18.4; DB 10; Length 582;
Best Local Similarity 95.0%; Pred. No. 8.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN

```

```

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 428 CAAGTACATGCAGAAACAGG 409

RESULT 4
LOCUS BP159851 591 bp mRNA linear EST 30-DEC-2003
DEFINITION BP159851 full-length enriched swine cDNA library, adult thymus Sus
scrofa cDNA clone THY010108A05 5', mRNA sequence.
ACCESSION BP159851
VERSION BP159851.1 GI:40409324
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.

REFERENCE 1 (bases 1 to 591)
AUTHORS Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H.,
Okumura,N., Hamasima,N. and Awata,T.
TITLE PEDE (Pig EST Data Explorer): construction of a database for ESTs
derived from porcine full-length cDNA libraries
JOURNAL Nucleic Acids Res. 32 (1), D484-D488 (2004)
PUBMED 14681463
COMMENT Contact: Hirohide Uenishi
Animal Genome Laboratory, Genome Research Department
National Institute of Agrobiological Sciences
2 Ikendodai, Tsukuba, Ibaraki 305-8602, Japan
Tel: +81-29-838-8627
Fax: +81-29-838-8627
Email: huenishi@affrc.go.jp
EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA
library
Vector sequences were eliminated by RepeatMasker version 2002/07/13
and crossmatch version 0.990319
Low quality bases were trimmed based on the quality values.

FEATURES
    source
    Location/Qualifiers
        1..591
        /organism="Sus scrofa"
        /mol_type="mRNA"
        /db_xref="taxon:9823"
        /clone="THY010108A05"
        /tissue_type="thymus"
        /dev_stage="adult"
        /clone_lib="full-length enriched swine cDNA library, adult
        thymus"

ORIGIN
Query Match 92.0%; Score 18.4; DB 3; Length 591;
Best Local Similarity 95.0%; Pred. No. 8.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 128 CAGGTAATGCAGAAACAGG 147

RESULT 5
LOCUS BP1770737 599 bp mRNA linear EST 25-SEP-2001
DEFINITION BP1770737 599 bp mRNA linear EST 25-SEP-2001
mRNA sequence.
ACCESSION BP1770737
VERSION BP1770737.1 GI:15762315
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini,

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 577 CAAGTAAATGCAGAAACAGG 596

RESULT 6
LOCUS BP1770737 599 bp mRNA linear GSS 04-MAY-1999
DEFINITION BP1770737 599 bp mRNA linear GSS 04-MAY-1999
genomic survey sequence.
ACCESSION BP1770737
VERSION BP1770737.1 GI:4044133
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini,
Hominidae; Homo.
REFERENCE 1 (bases 1 to 484)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Other_GSSs: RPC111-10214.TJ
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC

```

```

REFERENCE 1 (bases 1 to 599)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1528 row: a column: 23
High quality sequence start: 5
High quality sequence stop: 563.
Location/Qualifiers
    1..599
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5210398"
    /lab_host="DH10B"
    /clone_lib="NIH_MGC_122"
    /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
    Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
    anonymous pool of 24 week female lung, 16 week female
    spleen, and 20-22 week male spleens. Library is oligo-dT
    primed and directionally cloned (EcoRV site is destroyed
    upon cloning). Average insert size 1.4 kb, insert size
    range 1-3 kb. Library is normalized and enriched for
    full-length clones and was constructed by C. Gruber
    (Invitrogen). Research Genetics tracking code 026. Note:
    this is a NIH_MGC Library."

ORIGIN
Query Match 92.0%; Score 18.4; DB 3; Length 599;
Best Local Similarity 95.0%; Pred. No. 8.4e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 577 CAAGTAAATGCAGAAACAGG 596

RESULT 6
LOCUS BP1770737 599 bp mRNA linear GSS 04-MAY-1999
DEFINITION BP1770737 599 bp mRNA linear GSS 04-MAY-1999
genomic survey sequence.
ACCESSION BP1770737
VERSION BP1770737.1 GI:4044133
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini,
Hominidae; Homo.
REFERENCE 1 (bases 1 to 484)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C.
TITLE Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Other_GSSs: RPC111-10214.TJ
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are derived from the human BAC library RPCI-11. For BAC

```

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page: [http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html)  
 Seq primer: 17  
 Class: BAC ends.

#### FEATURES

Location/Qualifiers  
 1..484  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GB:7538979"  
 /db\_xref="taxon:9606"  
 /clone="RPC1-11-10214"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /clone\_lib="RPC1-11"  
 /note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI; RPC111 Human Male BAC Library"

#### ORIGIN

Query Match 90.0%; Score 18; DB 9; Length 484;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 2 AAGTAAATGCAGAAACAG 19  
 |||||  
 Db 16 AAGTAAATGCAGAAACAG 33

#### RESULT 7

CK593526  
 LOCUS  
 DEFINITION CK593526 183 bp mRNA linear EST 27-JAN-2004  
 tad32a01.y2 Hydra EST -Kiel 1 Hydra magnipapillata cDNA 5', mRNA

#### ACCESSION

CK593526

#### VERSION

CK593526.1 GI:41022181

#### KEYWORDS

EST.

#### SOURCE

Hydra magnipapillata

#### ORGANISM

Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydrozoa; Anthomedusae;

#### REFERENCE

1 (bases 1 to 183)  
 Bode,H., Blumberg,B., Steele,R., Wigge,P., Gee,L., Nguyen,Q.,  
 Martinez,D., Kibler,D., Hampson,S., Clifton,S., Pape,D., Marra,M.,  
 Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B., Bowers,Y.,  
 Gibbons,M., Ritter,E., Bennett,J., Ronko,I., Teagareishvili,R.,  
 Maguire,L., Kennedy,S., Waterston,R. and Wilson,R.

#### TITLE

WashU Hydra EST Project  
 Unpublished (2002)  
 Contact: Hans Bode  
 WashU Hydra EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)

Library was constructed by Konstantin Khalturin, Zoologisches  
 Institut, Univ. Kiel, Germany Library materials provided by Thomas  
 Bosch, Zoologisches Institut, CAU Kiel, Germany DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Hans Bode ([hzbode@uci.edu](mailto:hzbode@uci.edu))  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 183.

#### FEATURES

Location/Qualifiers  
 1..183  
 /organism="Hydra magnipapillata"  
 /mol\_type="mRNA"  
 /strain="105"  
 /db\_xref="taxon:6085"  
 /lab\_host="DH5a"  
 /clone\_lib="Hydra EST -Kiel 1"  
 /note="Vector: pSPORT1; Site.1: Not I; Site.2: Sal I;

pSPORT 1 Vector is ampicillin resistant, M13 reverse  
 primer was used by us for sequencing of 5' parts of  
 inserts; 3' parts of cDNAs contain long polyA tracks which  
 makes sequencing from 3' direction complicated"

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 7; Length 183;  
 Best Local Similarity 94.7%; Pred. No. 2.1e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 CAAGTAAATGCAGAAACAG 19  
 |||||  
 Db 142 CAAGTAAATGCAGAAACAG 160

#### RESULT 8

AZ464666

#### LOCUS

DEFINITION AZ464666 197 bp DNA linear GSS 04-OCT-2000  
 lmo274H09F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0274H09 F, genomic survey sequence.

#### ACCESSION

AZ464666

#### VERSION

AZ464666.1 GI:10622791

#### KEYWORDS

GSS.

#### SOURCE

Mus musculus (house mouse)

#### ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muroidea; Muridae; Murinae; Mus.

#### REFERENCE

1 (bases 1 to 197)  
 Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
 Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
 Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
 Niederhausern,A. and Wright,D., Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah Genome Center  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177

#### AUTHORS

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0274 row: H column: 09  
 Seq primer: CGTTGTAACGACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 197.

#### TITLE

Location/Qualifiers

#### JOURNAL

COMMENT

#### FEATURES

source

1..197  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0274H09"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male) was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (<http://www.jax.org/resources/documents/dnares/>). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adapted DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of PWD42 (gi|4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated

with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 197;  
Best Local Similarity 94.7%; Pred. No. 2.2e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAACAG 20  
|||||  
DB 84 AAGTAAATGCAGAACAG 102  
|||||

## RESULT 9

AL649495/c  
LOCUS  
DEFINITION  
AL649495 274 bp mRNA linear EST 20-NOV-2003  
Xenopus tropicalis cdna clone Tgas049j03 5',  
mRNA sequence.

ACCESSION  
AL649495

VERSION  
AL649495.2 GI:38463578

KEYWORDS  
EST.

SOURCE  
Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;  
Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 274)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

On Dec 13, 2001 this sequence version replaced gi:17658691.

Contact: Huckle E

Sahger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from sug of poly A+ RNA from stages 10-13  
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with  
EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI

Host: Escherichia coli XL1-blue

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE ID: Tgas049j03.plkSP6

Sequencing primer: SP6.

Location/Qualifiers

source

1. .274

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="Tgas049j03"

/dev\_stage="gastrula (stages 10-5-12 mixed)"

/lab\_host="Escherichia coli XL1-blue"

/clone\_lib="XGC-gastrula"

/note="vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA  
was oligo dt primed from sug of poly A+ RNA from stages  
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated  
into pCS107 with EcoRI at the 5' end and NotI at the 3'  
end."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 274;  
Best Local Similarity 94.7%; Pred. No. 2.3e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAG 19  
|||||

DB 21 CAAGTAAATGCAGAACAG 3  
|||||

## RESULT 10

AL679022/c

LOCUS

DEFINITION

AL679022 336 bp mRNA linear EST 10-NOV-2003

Xenopus tropicalis cdna clone TNeu055a09 5',

mRNA sequence.

ACCESSION

AL679022

VERSION

AL679022.2 GI:38251722

KEYWORDS

EST.

SOURCE

Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 336)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

On Mar 18, 2002 this sequence version replaced gi:19535396.

Contact: Taylor R

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from sug of poly A+ RNA from neurula.

EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the

5' end and NotI at the 3' end.

Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI

Host: Escherichia coli DH10B

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE ID: TNeu055a09.plcSP6

Sequencing primer: SP6.

Location/Qualifiers

source

1. .336

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TNeu055a09"

/dev\_stage="neurula"

/lab\_host="Escherichia coli DH10B"

/clone\_lib="XGC-neurula"

/note="vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA

was oligo dt primed from sug of poly A+ RNA from neurula.

EcoRI-NotI cut cDNA was then ligated into pCS107 with

EcoRI at the 5' end and NotI at the 3' end."

ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 336;

Best Local Similarity 94.7%; Pred. No. 2.3e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAG 19  
|||||

DB 104 CAAGTAAATGCAGAACAG 86  
|||||

RESULT 11

EX689357

LOCUS

DEFINITION

EX689357 350 bp mRNA linear EST 14-NOV-2003

Xenopus tropicalis cdna clone TNeu055a09 3',

mRNA sequence.

ACCESSION

EX689357

VERSION

EX689357.1 GI:38338477

KEYWORDS

EST.

SOURCE

Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 350)

Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.

Sanger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

```

COMMENT
Contact: Croning MDR
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS SEQUENCE ID: TNeu055a09.q1kt7
Sequencing primer: T7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli DH10B.
Location/Qualifiers
1..350
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TNeu055a09"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/clone_lib="XGC-neurula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

FEATURES
source
Query Match 87.0%; Score 17.4; DB 5; Length 350;
Best Local Similarity 94.7%; Pred. No. 2.3e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
|||||
Db 249 CAAGTAAATGCAGAAACAG 267

RESULT 12
B5666884
LOCUS 150514 MARC 4BOV Bos taurus cDNA 5', mRNA linear EST 25-APR-2001
ACCESSION B5666884
VERSION B5666884.1 GI:10027412
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
REFERENCE
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C.,
Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A.,
Chitko-McKown,C.G., Perteira,G., Holt,I., Karamycheva,S., Liang,F.,
Quackenbush,J. and Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
PUBMED 11282978
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 57 row: P column: 8

Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..412
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 4BOV"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from day 20 and day 40
embryos."

FEATURES
source
Query Match 87.0%; Score 17.4; DB 2; Length 412;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20
|||||
Db 79 AAGTAAATGCAGAAACAGG 97

RESULT 13
AW035216/c
LOCUS AW035216/c tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION AW035216
ACCESSION AW035216
VERSION AW035216.1 GI:5893972
KEYWORDS EST.
SOURCE Lycopersicon esculentum (Solanum lycopersicum)
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,
Holt,I.E., Liang,F., Opton,J., Craven,M.B., Bowman,C.B., Ahn,S.,
Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and
Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
Location/Qualifiers
1..429
/organism="Lycopersicon esculentum"
/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC35C9"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/clone_lib="tomato callus, TAMU"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"

FEATURES
source
Query Match 87.0%; Score 17.4; DB 1; Length 429;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19

```



```

Db      153 CATGTAATTCAGAAACAG 135
|||||
B89397 429 bp DNA linear GSS 09-APR-1999
LOCUS   RPCI11-24119.TPB RPCI-11 Homo sapiens genomic clone RPCI-11-24119,
DEFINITION genomic survey sequence.
ACCESSION B89397
VERSION   B89397.1 GI:2971070
KEYWORDS  GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
          Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
          Venter,J.C.
          1 (bases 1 to 429)
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.
          Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
          Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
          Venter,J.C.
          Use of BAC End Sequences for Sequence-Ready Map Building (1998)
          Unpublished (1998)
          Other_GSSs: RPCI11-24119.TV
          Contact: Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: mdadam@tigr.org
          Clones are derived from the human BAC library RPCI-11. For BAC
          library availability, please contact Pieter de Jong
          (pieter@ed.jong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
          Research Genetics (info@resgen.com). BAC end search page:
          http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
          Seq primer: SP6
          Class: BAC ends.
          Location/Qualifiers
          1..429
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="GDB:7509042"
             /db_xref="taxon:9606"
             /clone="RPCI-11-24119"
             /sex="Male"
             /cell_type="Lymphocytes"
             /clone_lib="RPCI-11"
             /notes="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
             RPCI11 Human Male BAC Library"

FEATURES
source
Query Match 87.0%; Score 17.4; DB 9; Length 429;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN
1 CAAGTAAATTCAGAAACAG 19
|||||
261 CAAGTAAATTCAGAAACAG 243

RESULT 15
AL966115/c
LOCUS   AL966115 XGC-gastrula Xenopus tropicalis cDNA clone TGAs090o12 5',
DEFINITION mRNA sequence.
ACCESSION AL966115
VERSION   AL966115.2 GI:39020253
KEYWORDS  EST.
SOURCE    Xenopus tropicalis (western clawed frog)
ORGANISM  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Query Match 87.0%; Score 17.4; DB 9; Length 429;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN
1 CAAGTAAATTCAGAAACAG 19
|||||
261 CAAGTAAATTCAGAAACAG 243

RESULT 14
B89397/c
LOCUS   RPCI11-24119.TPB RPCI-11 Homo sapiens genomic clone RPCI-11-24119,
DEFINITION genomic survey sequence.
ACCESSION B89397
VERSION   B89397.1 GI:2971070
KEYWORDS  GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
          Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
          Venter,J.C.
          1 (bases 1 to 429)
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.
          Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
          Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and
          Venter,J.C.
          Use of BAC End Sequences for Sequence-Ready Map Building (1998)
          Unpublished (1998)
          Other_GSSs: RPCI11-24119.TV
          Contact: Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: mdadam@tigr.org
          Clones are derived from the human BAC library RPCI-11. For BAC
          library availability, please contact Pieter de Jong
          (pieter@ed.jong.med.buffalo.edu). Clones may be purchased from
          BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
          Research Genetics (info@resgen.com). BAC end search page:
          http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
          Seq primer: SP6
          Class: BAC ends.
          Location/Qualifiers
          1..429
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="GDB:7509042"
             /db_xref="taxon:9606"
             /clone="RPCI-11-24119"
             /sex="Male"
             /cell_type="Lymphocytes"
             /clone_lib="RPCI-11"
             /notes="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
             RPCI11 Human Male BAC Library"

FEATURES
source
Query Match 87.0%; Score 17.4; DB 1; Length 433;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

ORIGIN
1 CAAGTAAATTCAGAAACAG 19
|||||
383 CAAGTAAATTCAGAAACAG 365

RESULT 16
AQ636375/c
LOCUS   AQ636375.1 GI:5099010
DEFINITION GSS.
ACCESSION AQ636375
VERSION   AQ636375.1 GI:5099010
KEYWORDS  GSS.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
REFERENCE
AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
          Venter,J.C.
          1 (bases 1 to 438)
          Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
          Homnidae; Homo.
          Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
          Venter,J.C.
          Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
          Map Building (1997)
          Unpublished (1997)
          Contact: Shaying Zhao, William Nierman, Mark Adams
          Department of Eukaryotic Genomics
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: hbeetigr.org
          Clones are derived from the human BAC library RPCI-11. For BAC

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 433)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
Unpublished (2003)
On Nov 27, 2002 this sequence version replaced gi:25789710.
Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TGAs090o12.plkasp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers
1..433
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGAs090o12"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/notes="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

```

library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs ([info@regen.com](mailto:info@regen.com)). BAC end search page: [http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html).  
Seq primer: SP6  
Class: BAC ends.

#### FEATURES

source  
1. .438  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GB:7683539"  
/db\_xref="taxon:9606"  
/clone="RPC1-11-478P12"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_lib="RPC1-11"  
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;  
RPC111 Human Male BAC Library"

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 438;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 AAGTAAATGCAGAAACAGG 20  
DB 142 AAGTAAATACAGAAACAGG 124

#### RESULT 17

BP090630/c  
LOCUS BP090630 472 bp mRNA linear EST 16-MAR-2005  
DEFINITION clone LC16CC06 5', mRNA sequence.  
ACCESSION BP090630  
VERSION BP090630.1 GI:58253101  
KEYWORDS EST.

SOURCE Lycopersicon esculentum (Solanum lycopersicum)  
ORGANISM Lycopersicon esculentum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;  
asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

#### REFERENCE

1 (bases 1 to 472)  
Yamamoto, N., Tsugane, T., Watanabe, M., Yano, K., Maeda, F., Kuwata, C., Moez, T., Nishimura, S. and Shibata, D.  
Expressed sequence tags from the laboratory-grown miniature tomato (Lycopersicon esculentum) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars

#### JOURNAL

COMMENT Unpublished (2005)  
Contact: Daisuke Shibata  
Kazusa DNA Research Institute;  
2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan  
Tel: 81-438-52-3947  
Fax: 81-438-52-3948  
Email: shibata@kazusa.or.jp

This clone was obtained at our laboratory.  
Please visit our web site

URL: <http://www.kazusa.or.jp/jmol/microtom/indexj.html> (in Japanese)  
URL: <http://www.kazusa.or.jp/jmol/microtom/indexj.html> (in English).

#### FEATURES

source  
1. .472  
/organism="Lycopersicon esculentum"  
/mol\_type="mRNA"  
/cultivar="Micro-Tom"  
/db\_xref="taxon:4081"  
/clone="LC16CC06"  
/tissue\_type="leaf"  
/clone\_lib="Lycopersicon esculentum leaf"

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 3; Length 472;

Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAG 19  
DB 136 CATGTAATGCAGAAACAG 118

#### RESULT 18

BP090630/c  
LOCUS BP090630 473 bp mRNA linear EST 31-AUG-2004  
DEFINITION BP090630 Yutaka Satou unpublished cDNA library (csga) Ciona savignyi cDNA clone csga089j14 3', mRNA sequence.

ACCESSION BP090630  
VERSION BP090630.1 GI:51734858  
KEYWORDS EST.

#### SOURCE

Ciona savignyi  
ORGANISM Ciona savignyi  
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
Phlebobranchia; Cionidae; Ciona.

#### REFERENCE

1 (bases 1 to 473)  
Vinson, J.P., Jaffe, D.B., O'Neill, K., Karlsson, E.K., Stange-Thomann, N., Anderson, S., Mesirov, J.P., Satoh, N., Satou, Y., Nuebaum, C., Birren, B., Galagan, J.E. and Lander, E.S.  
Assembly of polymorphic genomes: Algorithms and application to

#### TITLE

Ciona savignyi

#### JOURNAL

PUBMED 16077012

#### COMMENT

Contact: Yutaka Satou  
Department of Zoology  
Kyoto University  
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
Tel: 81-75-753-4095  
Fax: 81-75-705-1113  
Email: yutaka@ascidian.zool.kyoto-u.ac.jp.

#### FEATURES

source  
1. 473  
/organism="Ciona savignyi"  
/mol\_type="mRNA"  
/db\_xref="taxon:51511"  
/clone="csga089j14"  
/dev\_stage="gastrula stage"  
/clone\_lib="Yutaka Satou unpublished cDNA library (csga)"

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 5; Length 473;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

#### QY

2 AAGTAAATGCAGAAACAGG 20

DB 460 AAGTAAATGCAGAAACAGG 442

#### RESULT 19

BP090630/c  
LOCUS BP090630 478 bp mRNA linear EST 28-NOV-2000  
DEFINITION UI-R-CA1-blif-e-02-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone UI-R-CA1-blif-e-02-0-UI 3', mRNA sequence.

#### ACCESSION

BP090630

#### VERSION

BP090630.1 GI:11397364

#### KEYWORDS

EST.

#### SOURCE

Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridea; Muridae; Murinae; Rattus.

#### REFERENCE

1 (bases 1 to 478)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene discovery

#### JOURNAL

PUBMED 8889548

## COMMENT

Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@iowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized pons library cDNA Library Preparation: M.B. Soares Lab  
Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 21-150, >ORR1B#LTR/MaLR  
Seq primer: M13 Forward  
POLYA=Yes.

## FEATURES

Location/Qualifiers

source

```

1..478
/organism="Rattus norvegicus"
/mol_type="mrna"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-CA1-b1f-e-02-0-UI"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-CA1"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CA1
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla,
pons, midbrain, cerebral cortex, corpus striatum, testis,
and hippocampus. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subtraction
has been previously described in (Bonaldo, Lennon and
Soares, Genome Research 6:791-806, 1996)
TAG TISSUE=pons
TAG_LIB=UI-R-CA1
TAG_SEQ=AGCAGC"

```

## ORIGIN

```

Query Match      87.0%; Score 17.4; DB 2; Length 478;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 2 AAGTAAATGCAGAAACAGG 20

```

|||||
157 AAGTAAATGCAGATCAGG 139

```

RESULT 20

BP906911/c

LOCUS

```

DEFINITION BP906911 Lycopersicon esculentum leaf Lycopersicon esculentum cDNA
clone LC06AA08 5', mRNA sequence.

```

ACCESSION

BP906911

VERSION

BP906911.1

KEYWORDS

EST.

SOURCE

ORGANISM

Lycopersicon esculentum

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;

asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.

1 (bases 1 to 478)

Yamamoto, N., Tsugane, T., Watanabe, M., Yano, K., Maeda, F., Kuwata, C.,

Moez, T., Nishimura, S. and Shibata, D.

Expressed sequence tags from the laboratory-grown miniature tomato

(Lycopersicon esculentum) cultivar Micro-Tom and mining for single

nucleotide polymorphisms and insertions/deletions in tomato

cultivars

Unpublished (2005)

Contact: Daisuke Shibata

Kazusa DNA Research Institute;

2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan

Tel: 81-438-52-3947

Fax: 81-438-52-3948

Email: shibata@kazusa.or.jp

This clone was obtained at our laboratory.

Please visit our web site

URL: <http://www.kazusa.or.jp/jcol/microtom/indexj.html> (in Japanese)URL: <http://www.kazusa.or.jp/jcol/microtom/indexj.html> (in English).

## FEATURES

source

```

1..478
/organism="Lycopersicon esculentum"
/mol_type="mrna"
/cultivar="Micro-Tom"
/db_xref="taxon:4081"
/clone="LC06AA08"
/tissue_type="leaf"
/clone_lib="Lycopersicon esculentum leaf"

```

## ORIGIN

```

Query Match      87.0%; Score 17.4; DB 3; Length 478;
Best Local Similarity 94.7%; Pred. No. 2.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 CAAGTAAATGCAGAAACAG 19

```

|||||
143 CATGTAATGCAGAAACAG 125

```

RESULT 21

AZ961721/c

LOCUS

```

DEFINITION 2M0230G13F Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0230G13 F, genomic survey sequence.

```

ACCESSION

AZ961721

VERSION

AZ961721.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Mus musculus

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Murioidea; Muridae; Murinae; Mus.

1 (bases 1 to 480)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von

Niederhausern, A. and Wright, D., Weies, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: [ddunn@genetics.utah.edu](mailto:ddunn@genetics.utah.edu)

Insert Length: 10000 Std Error: 0.00

Plate: 0230 row: G column: 13

Seq primer: CGTTGTAACGACGCGCAGT

Class: plasmid ends

High quality sequence stop: 480.

Location/Qualifiers

source

1..480

/organism="Mus musculus"

/mol\_type="genomic DNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="UUGC2M0230G13"

/sex="Female"

/lab\_host="E. coli strain XL10-Gold, Tl-resistant, F-"

/clone\_lib="Mouse 10kb plasmid UUGC2M library"

/note="Vector: pWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi|4732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 480;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19

Db 384 CAAGTAATGCAGAAACAG 366

## RESULT 22

AL782393/c

LOCUS

DEFINITION AL782393 XGC-gastrula Xenopus tropicalis cDNA clone TGas074012 5', mRNA sequence.

ACCESSION AL782393

VERSION AL782393.2 GI:38486081

SOURCE Xenopus tropicalis

ORGANISM Xenopus tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 489)

AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)

JOURNAL Unpublished (2003)

COMMENT On Jun 25, 2002 this sequence version replaced gi:21568097.

Contact: Taylor R

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13

gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with

EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli XL1-blue

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TGas074012.plkSP6

Sequencing primer: SP6.

Location/Qualifiers

1. 489

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TGas074012"

/dev\_stage="gastrula (stages 10.5-12 mixed)"

/lab\_host="Escherichia coli XL1-blue"

/clone\_lib="XGC-gastrula"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA

was oligo dt primed from Sug of poly A+ RNA from stages

10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 1; Length 489;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCAGAAACAG 19

Db 384 CAAGTAATGCAGAAACAG 366

## RESULT 23

CO335661/c

LOCUS

DEFINITION EN10932.5prime Exelixis FlyTag MN08 Bluescript Drosophila melanogaster cDNA clone EN10932 5, mRNA sequence.

ACCESSION CO335661

VERSION CO335661.1 GI:49395936

KEYWORDS EST.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 490)

AUTHORS Nakanishi,M., Muzong,C., Peterson,E., Laufer,A., Leung,W., Platt,D.

and Swimmer,C.

EXelixis FlyTag EST Project MN08 Library

Unpublished (2004)

COMMENT BDGP

Lawrence Berkeley National Lab

One Cyclotron Rd, Berkeley, CA 94720, USA

Fax: 510 486 6798

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

Plate: EN109 row: C column: 8

High quality sequence stop: 394.

Location/Qualifiers

1. 490

/organism="Drosophila melanogaster"

/mol\_type="mRNA"

/db\_xref="taxon:7227"

/clone="EN10932"

/cell\_line="mbn2"

/clone\_lib="Exelixis FlyTag MN08 Bluescript"

/note="Vector: pBluescript; Site 1: NotI; Site 2: XhoI;

oligoDT primed from LPS induced mbn2 cell line."

## ORIGIN

Query Match 87.0%; Score 17.4; DB 7; Length 490;  
Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAATGCAGAAACAGG 20

Db 374 AGGTAATGCAGAAACAGG 356

## RESULT 24

AQ013410/c

LOCUS

DEFINITION RPC11-24119.TKBR RPCI-11 Homo sapiens genomic clone RPCI-11-24119, genomic survey sequence.

ACCESSION AQ013410

VERSION AQ013410.1 GI:3185975

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.  
 1 (bases 1 to 491)  
**REFERENCE**  
**AUTHORS**  
 Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Baas, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and Venter, J.C.  
**TITLE**  
**JOURNAL**  
**COMMENT**  
 Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
 Unpublished (1998)  
 Other GSSs: RPC111-24119.TPB RPC111-24119.TV  
**CONTACT**  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9710 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genetics ([info@resgen.com](mailto:info@resgen.com)). BAC end search page: [http://www.tigr.org/tcdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tcdb/humgen/bac_end_search/bac_end_search.html)  
 Class: BAC ends.  
**FEATURES**  
 source  
 1..491  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GDB:7509042"  
 /db\_xref="taxon:9606"  
 /clone="RPC1-11-24119"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /clone\_lib="RPC1-11"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPC111 Human Male BAC Library"  
**ORIGIN**  
 Query Match 87.0%; Score 17.4; DB 9; Length 491;  
 Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACAG 19  
 |||||  
 Db 242 CAAGTAAATTCAGAAACAG 224  
 |||||  
**RESULT 25**  
**AL672590/c**  
**LOCUS**  
**DEFINITION**  
 AL672590 XGC-gastrula Xenopus tropicalis cDNA clone TGas054d02 5', mRNA sequence.  
**ACCESSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**  
 Xenopus tropicalis (western clawed frog)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Silurana.  
**REFERENCE**  
**AUTHORS**  
 Croning, M.D.R., Ashurst, J.L., Taylor, R., Zorn, A.M. and Rogers, J.  
**TITLE**  
**JOURNAL**  
**COMMENT**  
 Sanger Xenopus tropicalis EST project 2001 (11\_2003)  
 On Mar 18, 2002 this sequence version replaced gi:19528946.  
 Contact: Taylor R  
 Sanger Institute  
 Hinxton, Cambridgeshire, CB10 1SA, UK  
 Email: trop@sanger.ac.uk  
 This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
 cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end. Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI

Host: Escherichia coli XL1-blue  
 Sanger Xenopus tropicalis EST project 2001  
 TROPICALIS SEQUENCE ID: TGas054d02.p1cSP6  
 Sequencing primer: SP6.  
**FEATURES**  
 source  
 1..505  
 /organism="Xenopus tropicalis"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:8364"  
 /clone="TGas054d02"  
 /dev\_stage="gastrula (stages 10-5-12 mixed)"  
 /lab\_host="Escherichia coli XL1-blue"  
 /clone\_lib="XGC-gastrula"  
 /note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."  
**ORIGIN**  
 Query Match 87.0%; Score 17.4; DB 1; Length 505;  
 Best Local Similarity 94.7%; Pred. No. 2.4e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 CAAGTAAATGCAGAAACAG 19  
 |||||  
 Db 408 CAAGTAAATGCAGAAACAG 390  
 |||||  
**RESULT 26**  
**BP908357/c**  
**LOCUS**  
**DEFINITION**  
 BP908357 Lycopersicon esculentum leaf Lycopersicon esculentum cDNA clone LC11DB11 5', mRNA sequence.  
**ACCESSION**  
**VERSION**  
**KEYWORDS**  
**SOURCE**  
**ORGANISM**  
 Lycopersicon esculentum (Solanum lycopersicum)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; asterids; lamids; Solanales; Solanales; Solanaceae; Solanum; Lycopersicon.  
**REFERENCE**  
**AUTHORS**  
 Yamamoto, N., Teugane, T., Watanabe, M., Yano, K., Maeda, F., Kuwata, C., Moez, T., Nishimura, S. and Shibata, D.  
**TITLE**  
 Expressed sequence tags from the laboratory-grown miniature tomato (Lycopersicon esculentum) cultivar Micro-Tom and mining for single nucleotide polymorphisms and insertions/deletions in tomato cultivars  
 Unpublished (2005)  
**JOURNAL**  
**COMMENT**  
 Contact: Daisuke Shibata  
 Kazusa DNA Research Institute;  
 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba, 292-0818, Japan  
 Tel: 81-438-52-3947  
 Fax: 81-438-52-3948  
 Email: shibata@kazusa.or.jp  
 This clone was obtained at our laboratory.  
 Please visit our web site  
 URL: <http://www.kazusa.or.jp/jsol/microtom/indexj.html> (in Japanese)  
 URL: <http://www.kazusa.or.jp/jsol/microtom/indexj.html> (in English).  
**FEATURES**  
 source  
 1..525  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultivar="Micro-Tom"  
 /db\_xref="taxon:4081"  
 /clone="LC11DB11"  
 /tissue\_type="leaf"  
 /clone\_lib="Lycopersicon esculentum leaf"  
**ORIGIN**  
 Query Match 87.0%; Score 17.4; DB 3; Length 525;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+03;

```

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAACTAAATGCAGAAACAG 19
||| ||||| ||||| |||||
Db 129 CATGTAATGCAGAAACAG 111

RESULT 27
AL655162/c 526 bp mRNA linear EST 20-NOV-2003
LOCUS AL655162.1 XGC-gastrula Xenopus tropicalis cDNA clone TGas026e19 5',
DEFINITION mRNA sequence.
ACCESSION AL655162
VERSION AL655162.2 GI:38463186
KEYWORDS EST.
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 526)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT On Dec 13, 2001 this sequence version replaced gi:17666907.
Contact: Huckle E
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site 1: EcoRI; Site 2: NotI
Host: Escherichia coli XL1-blue
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TGas026e19.pikSP6
Sequencing primer: SP6.
Location/Qualifiers
1..526
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TGas026e19"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

FEATURES
source
Query Match 87.0%; Score 17.4; DB 1; Length 526;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAACTAAATGCAGAAACAG 19
||| ||||| ||||| |||||
Db 315 CAACTAAATGCAGAAACAG 297

RESULT 28
CE293235 541 bp DNA linear GSS 26-SEP-2003
LOCUS CE293235 tigr-gss-dog-17000333878560 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE293235
VERSION CE293235.1 GI:36075319
KEYWORDS GSS.
SOURCE Canis familiaris (dog)

```

```

ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 541)
Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1..541
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 87.0%; Score 17.4; DB 10; Length 541;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 20
||||| ||||| ||||| |||||
Db 316 AAGTAAATGCAGAAACAG 334

RESULT 29
BP131868 550 bp mRNA linear EST 05-NOV-2004
LOCUS BP131868 MAT001 Nicotiana tabacum cDNA clone BY3970, mRNA sequence.
DEFINITION BP131868
ACCESSION BP131868
VERSION BP131868.1 GI:32874753
KEYWORDS EST.
SOURCE Nicotiana tabacum (common tobacco)
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
asterids; lamids; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 550)
Matsuoka,K., Demura,T., Galis,I., Horiguchi,T., Sasaki,M.,
Tashiro,G. and Fukuda,H.
A Comprehensive Gene Expression Analysis Toward the Understanding
of Growth and Differentiation of Tobacco BY-2 Cells
Plant Cell Physiol. 45 (9), 1280-1289 (2004)
15509851
Contact: Ken Matsuoka
Morphogenesis Research Group
RIKEN Plant Science Center
1-7-2 Suehirocho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9575
Fax: 81-45-503-9573
Email: by2@psc.riken.go.jp, URL:http://mrg.psc.riken.go.jp/strc/
The cDNA library was constructed from mRNA isolated from lag (9 h),
log (72 h) and stationary (7 days) old BY-2 cells.
Location/Qualifiers
1..550
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="Bright Yellow No.2"
/db_xref="taxon:4097"

```

```

/clone="BY3970"
/cell_line="BY-2"
/clone_lib="MAT001"
/notes="Vector: pGEM-T easy; primer: M13 forward; mRNA
obtained from lag, log and stationary phase cells"

ORGANISM
Bos taurus

REFERENCE
AUTHORS
Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.A.,
Bachman,S., Green,C., Wright,C.L., Campos,E.J., Benson,L.D.,
Edwards,J., Liu,L., Osogawa,K., Womack,J.E., de Jong,P.J. and
Lewin,H.A.

TITLE
A cattle-human comparative map built with cattle BAC-ends and human
genome sequence

JOURNAL
PUBMED
COMMENT
Genome Res. 13 (8), 1966-1972 (2003)
Other GSSs: CH240_114P20.TV
Contact: Harris Lewin
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBMC) by University of Illinois at Urbana
Champaign, USA with funds provided by grant No. AG202-34480-11828
from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
Genome Sequencing Initiative)
Plate: 114 row: P column: 20
Seq primer: SP6
Class: BAC ends

FEATURES
source
Location/Qualifiers
1..561
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_114P20"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/notes="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull 11 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Query Match 87.0%; Score 17.4; DB 9; Length 561;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAGG 20
|||||
Db 227 AAGTAAATGCTGAACAGG 209

RESULT 32
AZ596869
LOCUS
DEFINITION
1M0410119F Mouse 10kb plasmid UUGCIM library Mus musculus genomic
clone UUGCIM0410119 F, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Mus musculus (house mouse)

REFERENCE
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,

/clone="BY3970"
/cell_line="BY-2"
/clone_lib="MAT001"
/notes="Vector: pGEM-T easy; primer: M13 forward; mRNA
obtained from lag, log and stationary phase cells"

ORGANISM
Bos taurus

REFERENCE
AUTHORS
Larkin,D.M., Everts-van der Wind,A., Rebeiz,M., Schweitzer,P.A.,
Bachman,S., Green,C., Wright,C.L., Campos,E.J., Benson,L.D.,
Edwards,J., Liu,L., Osogawa,K., Womack,J.E., de Jong,P.J. and
Lewin,H.A.

TITLE
A cattle-human comparative map built with cattle BAC-ends and human
genome sequence

JOURNAL
PUBMED
COMMENT
Genome Res. 13 (8), 1966-1972 (2003)
Other GSSs: CH240_114P20.TV
Contact: Harris Lewin
Department of Animal Sciences
University of Illinois at Urbana Champaign
1201 W. Gregory Dr., Urbana, IL 61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.htm). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBMC) by University of Illinois at Urbana
Champaign, USA with funds provided by grant No. AG202-34480-11828
from USDA-CSREES and AG99-35205-8534 from USDA/NRI (Livestock
Genome Sequencing Initiative)
Plate: 114 row: P column: 20
Seq primer: SP6
Class: BAC ends

FEATURES
source
Location/Qualifiers
1..561
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="breed: Hereford"
/db_xref="taxon:9913"
/clone="CH240_114P20"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/notes="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
Hereford bull 11 Domino 99375; CHORI-240 Bovine BAC
library (Male) produced by Pieter de Jong"

ORIGIN
Query Match 87.0%; Score 17.4; DB 9; Length 561;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAGG 20
|||||
Db 227 AAGTAAATGCTGAACAGG 209

RESULT 32
AZ596869
LOCUS
DEFINITION
1M0410119F Mouse 10kb plasmid UUGCIM library Mus musculus genomic
clone UUGCIM0410119 F, genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus
Mus musculus (house mouse)

REFERENCE
AUTHORS
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,

```



Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
 Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
 Niederhausern,A. and Wright,D.,Weiss,R.  
 Mouse whole genome scaffolding with paired end reads from 10kb  
 plasmid inserts  
 JOURNAL  
 COMMENT  
 Unpublished (2000)  
 Contact: Robert B. Weiss  
 University of Utah  
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0410 row: I column: 19  
 Seq primer: CGTTGTAAACACGCGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 562.  
 Location/Qualifiers  
 FEATURES  
 source  
 1..562  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0410119"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, Tl-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PMP42nv; Purified genomic DNA from M.  
 musculus C57BL/6J (male); was obtained from the Jackson  
 Laboratory Mouse DNA Resource  
 (http://www.jax.org/resources/documents/dnares/). The DNA  
 was hydrodynamically sheared by repeated passage through a  
 0.005 inch orifice at constant velocity. The sheared DNA  
 was blunt end-repaired with T4 DNA polymerase and T4  
 polynucleotide kinase. Adaptor oligonucleotides were  
 ligated to the blunt ends in high molar excess. The  
 adaptor DNA was purified and size-selected for a 9.5 to  
 10.5 kb range using preparative agarose gel  
 electrophoresis. Vector DNA was prepared from a derivative  
 of pWD42 (gi|4732114|gb|AF129072.1), a copy-number  
 inducible derivative of plasmid R1. The vector was ligated  
 with adaptors complementary to the insert adaptors and  
 purified. The sheared, adaptor mouse DNA was annealed to  
 adaptor vector DNA, and transformed into  
 chemically-competent E. coli XL10-Gold (Stratagene) cells  
 and selected for ampicillin resistance."

ORIGIN  
 Query Match 87.0%; Score 17.4; DB 9; Length 562;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AAGTAAATGCAGAACAGG 20  
 ||||| ||||| ||||| |||||  
 Db 264 AAGTAAATACAGAACAGG 282

RESULT 33  
 AG929825/c  
 LOCUS  
 DEFINITION  
 Drosophila melanogaster DNA, clone: DME1-014C11.r.a, genomic  
 survey sequence.  
 ACCESSION  
 AG929825  
 VERSION  
 AG929825.1 GI:58451244  
 KEYWORDS  
 GSS.  
 SOURCE  
 Drosophila melanogaster (fruit fly)  
 ORGANISM  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

REFERENCE  
 1  
 AG929825  
 Drosophila melanogaster DNA, clone: DME1-014C11.r.a, genomic  
 survey sequence.  
 ACCESSION  
 AG929825  
 VERSION  
 AG929825.1 GI:58451244  
 KEYWORDS  
 GSS.  
 SOURCE  
 Drosophila melanogaster (fruit fly)  
 ORGANISM  
 Drosophila melanogaster  
 Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.

AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Hattori,M., Toyoda,A., Murakami,K., Kuroki,Y., Fujiyama,A.,  
 Toshio,T.K. and Sakaki,Y.  
 BAC end sequences of Library DME1  
 Unpublished  
 2 (bases 1 to 564)  
 Hattori,M.  
 Direct Submission  
 Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa, 230-0045, Japan  
 (E-mail:hattori@psc.riken.jp, URL:http://hgp.gsc.riken.jp/,  
 Tel:81-45-503-9111, Fax:81-45-503-9170)  
 Clones are derived from the BAC library DME1  
 For BAC library availability, please contact Masa-Toshi Yamamoto  
 (yamamoto@kit.jp).  
 Submitted (30-11-2004) by Masahira Hattori,  
 RIKEN, Genomic Sciences Center (GSC);  
 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail:hattori@psc.riken.jp,Tel:81-45-503-9111,  
 Fax:81-45-503-9170)  
 This work was done in collaboration with Yamamoto, M-T. Drosophila  
 Genetic Resource Center  
 Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan  
 Tel: 81-75-873-2660 FAX: 81-75-861-0881  
 PRIMERS  
 Sequencing : R  
 LIBRARY  
 Vector : pKS150  
 R.Site 1 : SacI  
 R.Site 2 : SacI  
 Location/Qualifiers  
 1..564  
 /organism="Drosophila melanogaster"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7227"  
 /clone="DME1-014C11.R.fa"  
 /clone\_lib="DME1 Drosophila BAC library"

ORIGIN  
 Query Match 87.0%; Score 17.4; DB 10; Length 564;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 AAGTAAATGCAGAACAGG 20  
 ||||| ||||| ||||| |||||  
 Db 240 AGGTAATGCAGAACAGG 222

RESULT 34  
 CE503452  
 LOCUS  
 DEFINITION  
 tigr-gss-dog-17000327346876 Dog Library Canis familiaris genomic,  
 genomic survey sequence.  
 ACCESSION  
 CE503452  
 VERSION  
 CE503452.1 GI:36820233  
 KEYWORDS  
 GSS.  
 SOURCE  
 Canis familiaris (dog)  
 ORGANISM  
 Canis familiaris  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;  
 Canis.

REFERENCE  
 1 (bases 1 to 568)  
 Kirkness,B.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,  
 Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and  
 Venter,J.C.  
 The dog genome: survey sequencing and comparative analysis  
 Science 301 (5641), 1898-1903 (2003)  
 14512627  
 TITLE  
 JOURNAL  
 PUBMED  
 COMMENT  
 Contact: Kirkness EF  
 The Institute for Genomic Research  
 Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
 Rockville, MD 20850, USA  
 Tel: 301-638-0200



Fax: 301-838-0208  
Email: ekirknes@tigr.org  
Class: shotgun.

# FEATURES

## Location/Qualifiers

1. .568  
/organism="Canis familiaris"  
/mol\_type="genomic DNA"  
/strain="Standard Poodle"  
/db\_xref="taxon:9615"  
/clone\_lib="Dog Library"  
/note="Site 1: BstXI; Libraries were prepared from peripheral blood"

# ORIGIN

Query Match 87.0%; Score 17.4; DB 10; Length 568;  
Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 AAGTAAATGCAGAAACAGG 20  
|||||  
Db 407 AAGTAAATGCAGAAACAGG 425  
|||||

# RESULT 35

CX997388 571 bp mRNA linear EST 09-FEB-2005  
JGI\_CAAQ1214.fwd NIH\_XGC\_trophHt1 Xenopus tropicalis cDNA clone  
IMAGE:7721993 5', mRNA sequence.

# LOCUS

## DEFINITION

CX997388

## ACCESSION

VERSION

## KEYWORDS

SOURCE

## ORGANISM

Xenopus tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1. (bases 1 to 571)

Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute

Unpublished (2004)

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Robert M. Grainger

cDNA Library Preparation: Bruce Blumberg Laboratory, University of

California, Irvine

DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>

Clone Distribution: I.M.A.G.E. Consortium/LLNL:

<http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone Id and the direction of sequencing. The suffix '.fwd'

indicates a forward sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Small Insert: Based upon one or more sequencing reads of this clone

where vector sequence was present at both ends, this clone has been

determined to contain a cDNA insert on the order of 600-1000 bases.

Plate: CAAQ 0013 row: k column: 15

High quality sequence stop: 571.

Location/Qualifiers

1. .571

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="N6 (Nigerian 6th generation inbred)"

/db\_xref="taxon:8364"

/clone="IMAGE:7721993"

/tissue\_type="Heart"

/dev\_stage="Adult"

/lab\_host="Electromax DH10B T1 Phage Resistant cells"

# FEATURES

## source

## Location/Qualifiers

1. .572

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="N6 (Nigerian 6th generation inbred)"

/db\_xref="taxon:8364"

/clone="IMAGE:7721993"

/dev\_stage="Adult"

/lab\_host="Electromax DH10B T1 Phage Resistant cells"

/lab\_host="Electromax DH10B T1 Phage Resistant cells"

/lab\_host="Electromax DH10B T1 Phage Resistant cells"

/lab\_host="Electromax DH10B T1 Phage Resistant cells"

/lab\_host="Electromax DH10B T1 Phage Resistant cells"

/clone\_lib="NIH\_XGC\_trophHt1"  
/note="Vector: pCS107; Site 1: EcoRI; Site 2: XhoI; The library was prepared from 5 ug of poly A+ RNA by oligo-dT priming  
(5'-ACTAGTCGGCGCTAGGCTCGAGTGTGTTTTTTTTTTTTTTT-3') and Stratascript reverse transcriptase. After ligation of EcoRI adapters (5'-AATTCGGCAGGAGG-3') followed by kinaasing adapters and by XhoI digestion, the cDNA was size selected by chromatography on Sepharose CL-2B columns and fractions containing cDNAs larger than 1000 bp were ligated into EcoRI/XhoI-digested pCS107. Reference for library construction: Current Genomics 4, 635-644. Library constructed by Michelle Tabb and Bruce Blumberg (Dept of Developmental and Cell Biology, University of California, Irvine)."

# ORIGIN

Query Match 87.0%; Score 17.4; DB 8; Length 571;  
Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 CAAATAATGCAGAAACAG 19  
|||||  
Db 220 CAAATAATGCAGAAACAG 238  
|||||

# RESULT 36

CX997387/c

# LOCUS

## DEFINITION

CX997387

## ACCESSION

VERSION

## KEYWORDS

SOURCE

## ORGANISM

Xenopus tropicalis (western clawed frog)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;

Xenopodinae; Xenopus; Silurana.

1. (bases 1 to 572)

Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,

Brokstein, P. and Lindquist, E.A.

DOE Joint Genome Institute

Unpublished (2004)

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Robert M. Grainger

cDNA Library Preparation: Bruce Blumberg Laboratory, University of

California, Irvine

DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>

Clone Distribution: I.M.A.G.E. Consortium/LLNL:

<http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone Id and the direction of sequencing. The suffix '.rev'

indicates a reverse sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Small Insert: Based upon one or more sequencing reads of this clone

where vector sequence was present at both ends, this clone has been

determined to contain a cDNA insert on the order of 600-1000 bases.

Plate: CAAQ 0013 row: k column: 15

High quality sequence stop: 572.

Location/Qualifiers

1. .572

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="N6 (Nigerian 6th generation inbred)"

/db\_xref="taxon:8364"

/clone="IMAGE:7721993"

/dev\_stage="Adult"

/lab\_host="Electromax DH10B T1 Phage Resistant cells"

/lab\_host="Electromax DH10B T1 Phage Resistant cells"



DEFINITION EC2CAA34AF06.b1 Xenopus tropicalis xthr plasmid library Xenopus tropicalis cDNA clone xthr34K11 3', mRNA sequence.

ACCESSION CN109679

VERSION CN109679.1 GI:45902375

KEYWORDS EST.

ORGANISM Xenopus tropicalis (western clawed frog)

SOURCE Xenopus tropicalis

REFERENCE 1 (bases 1 to 586)

AUTHORS Thuret,R., Fierro,A.C., Coen,L., Perron,M., Demeneix,B., Wegnez,M., Gyapay,G., Weissenbach,J., Wincker,P., Mazabraud,A. and Pollet,N.

TITLE Exploring the nervous system transcriptome in the model Xenopus tropicalis using EST analysis

JOURNAL Unpublished (2004)

COMMENT Contact: Pollet N  
Transgenese et Genetique des Amphibiens  
CNRS UMR 8080  
IBAC bat 447, Universite Paris Sud, Orsay, F-91405, France  
Tel: +33 169157272  
Fax: +33 169156816  
Email: Nicolas.Pollet@ibaic.u-psud.fr.

FEATURES  
source  
1..586  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/strain="Ivory Coast"  
/db\_xref="taxon:8364"  
/clone="xthr34K11"  
/tissue\_type="pool of heads and retinas from tailbud stages 25-35"  
/dev\_stage="stage 25-35"  
/lab\_host="E.coli DH10B"  
/clone\_lib="Xenopus tropicalis xthr plasmid library"  
/note="Vector: pCMVSPORT6 xthr; Site\_1: Sfil; Site\_2: Sfil; Xenopus tropicalis polyA+ RNA was obtained from pool of heads and retinas from tailbud stages 25-35 CDNAs were synthesized using the SMART system of CLONTECH and directionally cloned into pCMVSPORT6 xthr, a modified version of pCMVSPORT6 allowing directional cloning using asymmetric Sfil sites. For antisense RNA synthesis, use T7 promoter and for sense RNA use SP6 promoter. Library constructed by Drs. N. Pollet, M. Perron, M. Wegnez, A. Mazabraud (CNRS UMR 8080, Universite Paris Sud, Orsay, France)."

ORIGIN  
Query Match 87.0%; Score 17.4; DB 7; Length 586;  
Best Local Similarity 94.7%; Pred.No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCAAGTAAATGCAGAAACAG 19  
|||||  
Db 183 CCAAGTAAATGCAGAAACAG 201

RESULT 40  
BU919936/c

LOCUS BU919936.1 GI:24423788

DEFINITION 6042-56 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA, mRNA sequence.

ACCESSION BU919936

VERSION BU919936.1 GI:24423788

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 600)

AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W., White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.

TITLE Gene expression in the developing mouse retina by EST sequencing and microarray analysis

JOURNAL Nucleic Acids Res. 29 (24), 4983-4993 (2001)

COMMENT Contact: Klein WH  
Department of Biochemistry and Molecular Biology  
University of Texas M.D. Anderson Cancer Center  
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA  
Tel: 713 792 3646  
Fax: 713 790 0329.

FEATURES  
source  
1..600  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/tissue\_type="neural retina"  
/dev\_stage="embryonic day 14.5 post-fertilization"  
/clone\_lib="Mouse E14.5 retina lambda ZAP II Library"

ORIGIN  
Query Match 87.0%; Score 17.4; DB 5; Length 600;  
Best Local Similarity 94.7%; Pred.No. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCAAGTAAATGCAGAAACAG 19  
|||||  
Db 49 CCAAGTAAATGCAGAAACAG 31

RESULT 41  
AL644275/c

LOCUS AL644275.2 GI:38222278

DEFINITION 604 bp mRNA linear EST 07-NOV-2003  
mRNA sequence.

ACCESSION AL644275

VERSION AL644275.2

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM Xenopus tropicalis

REFERENCE 1 (bases 1 to 604)

AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

TITLE Sanger Xenopus tropicalis EST project 2001 (11\_2003)

JOURNAL Unpublished (2003)

COMMENT On Nov 7, 2001 this sequence version replaced gi:16796400.  
Contact: Huckle E  
Sanger Institute  
Hinxton, Cambridgeshire, CB10 1SA, UK  
Email: trop@sanger.ac.uk  
This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.  
cDNA was oligo dt primed from sug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.  
Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI  
Host: Escherichia coli DH10B  
Sanger Xenopus tropicalis EST project 2001  
TROPICALIS\_SEQUENCE\_ID: TNeu027117.plkSP6  
Sequencing primer: SP6.

FEATURES  
source  
1..604  
/organism="Xenopus tropicalis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8364"  
/clone="TNeu027117"  
/dev\_stage="neurula"  
/lab\_host="Escherichia coli DH10B"  
/clone\_lib="XGC-neurula"  
/note="Vector: pCS107; Site\_1: EcoRI; Site\_2: NotI; cDNA was oligo dt primed from sug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.">

EcoRI at the 5' end and NotI at the 3' end."

```

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 604;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAACAG 19
      |||||
Db      499 CAAGTAAATGAAGAACAG 481

RESULT 42
BG634601/c
LOCUS      604 bp      mRNA      linear      EST 23-APR-2001
DEFINITION  BG634601.1 Drosophila melanogaster adult testes pOTB7
              Drosophila melanogaster cDNA clone AT30859 5 similar to CytC;
              FBan006292 'cell cycle regulator' located on: 3L 74D2-74D3;
              04/09/2001, mRNA sequence.
ACCESSION  BG634601
VERSION    BG634601.1 GI:13762138
KEYWORDS   EST.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 604)
AUTHORS   Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Baxter,E.,
            Berman,B., Carlson,J., Champe,M., Chavez,C., Chew,M., Dorsett,V.,
            Farfan,D., Frise,E., George,R., Gonzalez,M., Guarin,H., Harris,N.,
            Li,P., Liao,G., Miranda,A., Misra,S., Mungall,C.J., Nunoo,J.,
            Pacieb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K., Yu,C.,
            Lewis,S.E., Celinker,S. and Rubin,G.M.
            BDGP/HMI AT Drosophila EST Project
            Unpublished (2000)
TITLE     Contact: Stapleton, M.
JOURNAL   BDGP
COMMENT   Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
            hit genomic AB003523: arm:3L [17391057,17674393]
            estimated-cyto:74D1-75A1: 04/09/2001
            Plate: AT.308 row: E column: 11
            High quality sequence stop: 603.
            Location/Qualifiers
            1..604
            /organism="Drosophila melanogaster"
            /mol_type="mRNA"
            /db_xref="taxon:7227"
            /clone="AT30859"
            /sex="male"
            /dev_stage="0-3 day old Ore-R males"
            /lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
            AT.121-AT.319: DHS-alpha Tona"
            /clone_lib="AT Drosophila melanogaster adult testes pOTB7"
            /note="Organ: ADULT testes; Vector: pOTB7; Site 1: EcoRI;
            Site 2: XhoI. The mRNA for the testis library was made
            from testes and seminal vesicles hand dissected from 0-3
            day old Ore-R males. RNA kindly provided by the lab of
            Margaret Fuller. Sized fractionated cDNAs were directly
            ligated into pOTB7. Plasmid cDNA library."

ORIGIN
Query Match      87.0%; Score 17.4; DB 2; Length 604;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAACAG 20
      |||||
Db      365 AGGTAAATGCAGAACAG 347

RESULT 43
CZ498169/c
LOCUS      607 bp      DNA      linear      GSS 11-MAY-2005
DEFINITION  OA_BB00142K02.f OA_BBa Oryza alta genomic clone OA_BB00142K02 5',
              genomic survey sequence.
ACCESSION  CZ498169
VERSION    CZ498169.1 GI:63218977
KEYWORDS   GSS.
SOURCE     Oryza alta
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE  1 (bases 1 to 607)
AUTHORS   SanMiguel,P., Westerman,R., Kim,H., Yu,Y., Wisotsaki,M., Yost,D.,
            Stum,D., Rao,K., Luo,M., Jetty,R., Kudrna,D., Muller,C.,
            Hatfield,J., Soderlund,C., Wing,R. and Jackson,S.A.
            OMAP (Oryza Map Alignment Project) - Purdue University
            Unpublished (2004)
TITLE     Contact: Scott A. Jackson
JOURNAL   Jackson Laboratory
COMMENT   Purdue University
            915 W. State St., West Lafayette, IN 47907, USA
            Tel: 7654963621
            Fax: 7654967255
            Email: sjackson@purdue.edu
            Basecalling by phred version 0.020425.c. This sequence was derived
            from the raw sequence read by clipping with Lucy version 1.198.
            Bases 31-637 of the raw sequence (length 992) were retained after
            clipping.
            Plate: 0142 row: K column: 02
            Seq primer: TAA TAC GAC TCA CTA TAG GG
            Class: BAC ends.
            Location/Qualifiers
            1..607
            /organism="Oryza alta"
            /mol_type="genomic DNA"
            /db_xref="taxon:52545"
            /clone="OA_BB00142K02"
            /tissue_type="young leaves"
            /lab_host="DH10B-T1 phage resistant"
            /clone_lib="OA_BBa"
            /note="Vector: pGIBAC1; Site_1: HindIII; Site_2: HindIII"

ORIGIN
Query Match      87.0%; Score 17.4; DB 10; Length 607;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAACAG 19
      |||||
Db      375 CAAGTAAATGCAGAACAG 357

RESULT 44
BX764191
LOCUS      610 bp      mRNA      linear      EST 10-DEC-2003
DEFINITION  BX764191 XGC-gastrula Xenopus tropicalis cDNA clone TCas090612 3',
              mRNA sequence.
ACCESSION  BX764191
VERSION    BX764191.1 GI:39671400
KEYWORDS   EST.
SOURCE     Xenopus tropicalis (western clawed frog)
ORGANISM   Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE  1 (bases 1 to 610)
AUTHORS   Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
            Sanger Xenopus tropicalis EST project 2001 (11_2003)
            Unpublished (2003)
JOURNAL   Contact: Croning MDR
COMMENT

```

```

Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TGA090012.q1kT7
Sequencing primer: 17
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers
1..610
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN
Query Match 87.0%; Score 17.4; DB 5; Length 610;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
|||||
Db 225 CAAGTAAATGCAGAAACAG 243

RESULT 45
AL793267/c
LOCUS AL793267 613 bp mRNA linear EST 13-NOV-2003
DEFINITION AL793267 XGC-neurula Xenopus tropicalis cDNA clone TNeu108a11 5',
mRNA sequence.
ACCESSION AL793267.2 GI:38315272
VERSION EST.
KEYWORDS Xenopus tropicalis (western clawed frog)
SOURCE Xenopus tropicalis
ORGANISM Xenopus tropicalis
REFERENCE 1 (bases 1 to 613)
AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT On Jun 25, 2002 this sequence version replaced gi:21578971.

Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the
5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli DH10B
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu108a11.plcSP6
Sequencing primer: SP6.
Location/Qualifiers
1..613
/organism="Xenopus tropicalis"

FEATURES
source
1..613
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN
Query Match 87.0%; Score 17.4; DB 1; Length 618;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
|||||
Db 434 CAAGTAAATGCAGAAACAG 416

RESULT 46
AL967548/c
LOCUS AL967548 618 bp mRNA linear EST 05-DEC-2003
DEFINITION AL967548 XGC-gastrula Xenopus tropicalis cDNA clone TGas131122 5',
mRNA sequence.
ACCESSION AL967548.2 GI:39023527
VERSION EST.
KEYWORDS Xenopus tropicalis (western clawed frog)
SOURCE Xenopus tropicalis
ORGANISM Xenopus tropicalis
REFERENCE 1 (bases 1 to 618)
AUTHORS Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.
TITLE Sanger Xenopus tropicalis EST project 2001 (11_2003)
JOURNAL Unpublished (2003)
COMMENT On Nov 27, 2002 this sequence version replaced gi:35791143.

Contact: Taylor R
Sanger Institute
Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TGas131122.plkSP6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13
gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end.
Vector: pCS107; Site_1: EcoRI; Site_2: NotI
Host: Escherichia coli XL1-blue.
Location/Qualifiers
1..618
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN
Query Match 87.0%; Score 17.4; DB 1; Length 618;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

FEATURES
source
1..618
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from Sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

```



```

/clone="TGas084j16"
/dev_stage="gastrula (stages 10.5-12 mixed)"
/lab_host="Escherichia coli XL1-blue"
/clone_lib="XGC-gastrula"
/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA
was oligo dt primed from sug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 620;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
    |||||
Db 408 CAAGTAAATGAAGAAACAG 390

RESULT 50
BQ521386/c
LOCUS      621 bp mRNA linear EST 10-JUN-2002
DEFINITION NISC n110901.y1 NTCHE XGC Emb7 Xenopus tropicalis cDNA clone
IMAGE:5336233 5', mRNA sequence.
ACCESSION  BQ521386
VERSION     BQ521386
KEYWORDS   EST.
SOURCE     Xenopus tropicalis (western clawed frog)
ORGANISM   Xenopus tropicalis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
            Xenopodinae; Xenopus; Silurana.
REFERENCE  1 (bases 1 to 621)
            NIH-XGC http://image.llnl.gov/image/html/xenopuslib.info.shtml.
            National Institute of Child Health and Human Development, National
            Cancer Institute, Xenopus Gene Collection
            Unpublished (2002)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs@mail.nih.gov
            cDNA Library Preparation:
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plate: L1AM11853 row: M column: 2
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             source
     source
     1..621
     /organism="Xenopus tropicalis"
     /mol_type="mRNA"
     /db_xref="taxon:8364"
     /clone="IMAGE:5336233"
     /tissue_type="tailbud"
     /dev_stage="embryo, stages 20-27"
     /lab_host="DH10B (phage-resistant)"
     /clone_lib="NTCHD XGC Emb7"
     /note="Vector: pCMV-SPORT6.1; Site 1: NotI; Site 2: EcoRV;
Cloned unidirectionally. Primer: Oligo dt. Average insert
size 2.1 kb. Constructed by Invitrogen. Note: This is a
Xenopus Gene Collection (XGC) library."

ORIGIN
Query Match      87.0%; Score 17.4; DB 5; Length 621;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
    |||||
Db 409 CAAGTAAATGAAGAAACAG 391

```

```

RESULT 51
AG931878/c
LOCUS      623 bp DNA linear GSS 01-FEB-2005
DEFINITION Drosophila melanogaster DNA, clone: DME1-017K16.F.fa, genomic
survey sequence.
ACCESSION  AG931878
VERSION     AG931878.1 GI:58453297
KEYWORDS   GSS.
SOURCE     Drosophila melanogaster (fruit fly)
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1
            Hattori,M., Toyoda,A., Murakami,K., Kuroki,Y., Fujiyama,A.,
            Toshio,T.K. and Sakaki,Y.
            BAC end sequences of Library DME1
            Unpublished
            2 (bases 1 to 623)
            Hattori,M.
            Direct Submission
            Submitted (19-JAN-2005) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
            Clones are derived from the BAC library DME1
            For BAC library availability, please contact Masa-Toshi Yamamoto
            (yamamoto@kit.jp)
            Submitted (30-11-2004) by Masahira Hattori,
            RIKEN, Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsuri-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.jp,Tel:81-45-503-9111,
            Fax:81-45-503-9170)
            This work was done in collaboration with Yamamoto, M-T. Drosophila
            Genetic Resource Center
            Saga Ippongi-cho, Ukyo-ku, Kyoto 616-8354, Japan
            Tel: 81-75-873-2660 FAX: 81-75-861-0881
            PRIMERS
            Sequencing : F
LIBRARY
Vector      : pKS150
R.Site 1    : SacI
R.Site 2    : SacI.
Location/Qualifiers
1..623
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="DME1-017K16.F.fa"
/clone_lib="DME1 Drosophila BAC library"

FEATURES             source
     source
     Query Match      87.0%; Score 17.4; DB 10; Length 623;
     Best Local Similarity 94.7%; Pred. No. 2.5e+03;
     Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 20
    |||||
Db 302 AGGTAATGCAGAAACAG 284

RESULT 52
CX466293/c
LOCUS      634 bp mRNA linear EST 08-AUG-2005
DEFINITION JGI XZG47538.fwd NIH XGC tropGae7 Xenopus tropicalis cDNA clone
IMAGE:7562072 5', mRNA sequence.
ACCESSION  CX466293
VERSION     CX466293.2 GI:71979250
KEYWORDS   EST.
SOURCE     Xenopus tropicalis (western clawed frog)
ORGANISM   Xenopus tropicalis

```



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.

# REFERENCE AUTHORS

1 (bases 1 to 634)

# TITLE JOURNAL

Richardson, P., Lucas, S., Rokhsar, D., Detter, J.C., Ng, D.C., Brokstein, P. and Lindquist, E.A.

# JOURNAL COMMENT

DOE Joint Genome Institute Xenopus tropicalis EST project

Unpublished (2004)

On Jan 10, 2005 this sequence version replaced gi:57352169.

Other ESTs: JGI.XZG47538.rev

Contact: Lindquist, E.A., Richardson, P.

DOE Joint Genome Institute

2800 Mitchell Drive, Walnut Creek, CA 94598, USA

Tel: 925 296 5600

Fax: 925 296 5710

Email: cdna@jgi-psf.org

Tissue Procurement: Richard M. Harland Laboratory, University of

California, Berkeley: <http://tropicalis.berkeley.edu/home>

CDNA Library Preparation: Richard M. Harland Laboratory, University

of California, Berkeley

DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>

Clone Distribution: I.M.A.G.E. Consortium/LLNL:

<http://image.llnl.gov>

Naming Conventions: EST name is generated by the concatenation of

the JGI Clone id and the direction of sequencing. The suffix '.fwd'

indicates a forward sequencing read of the insert. It does not

necessarily reflect the orientation of the insert.

Small Insert: Based upon one or more sequencing reads of this clone

where vector sequence was present at both ends, this clone has been

determined to contain a cDNA insert on the order of 600-1000 bases.

Plate: XZG 0493 row: d column: 6

High quality sequence stop: 634.

Location/Qualifiers

1. .634

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="IMAGE:7562072"

/tissue\_type="whole embryo"

/dev\_stage="Gastrula (st. 10.5-12.5)"

/lab\_host="E. coli XL1-Blue derivative, Stratagene

Electrofen-Blue"

/clone\_lib="NIH XGC tropGas7"

/note="Vector: pCS108; Site 1: SalI; Site 2: NotI;

Gastrula library constructed by Russell B. Fletcher in R.

Harland's lab using poly A RNA and oligo dt primers

(Invitrogen SuperScript plasmid System for cDNA Synthesis

and Cloning). SalI (5' end) -NotI (3' end) cDNA was

inserted into vector pCS108

(<http://mcb.berkeley.edu/labs/harland/pages/plasmids.html>)

"

Query Match 87.0%; Score 17.4; DB 8; Length 634;

Best Local Similarity 94.7%; Pred. No. 2.5e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCGAAGAACAG 19

|||||

Db 436 CAAGTAATGCGAAGAACAG 418

|||||

RESULT 53

CN109680/c

LOCUS

DEFINITION

EC2CA34AF06.g1 Xenopus tropicalis xthr plasmid library Xenopus

tropicalis cDNA clone xthr34kl1 5', mRNA sequence.

CN109680

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Xenopus; Silurana.

# REFERENCE AUTHORS

1 (bases 1 to 638)

# TITLE JOURNAL

Thuret, R., Fierro, A.C., Coen, L., Perron, M., Demeneix, B., Wegnez, M., Gvapat, G., Weissenbach, J., Wincker, P., Mazabraud, A. and Pollet, N.

# JOURNAL COMMENT

Exploring the nervous system transcriptome in the model Xenopus

tropicalis using EST analysis

Unpublished (2004)

Contact: Pollet N

Transgenese et Genetique des Amphibiens

CNRS UMR 8080

IBAC bat 447, Universite Paris Sud, Orsay, F-91405, France

Tel: +33 169157272

Fax: +33 169156816

Email: Nicolas.Pollet@ibaic.u-psud.fr.

Location/Qualifiers

1. .638

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="Ivory Coast"

/db\_xref="taxon:8364"

/clones="xthr34kl1"

/tissue\_type="pool of heads and retinas from tailbud

stages 25-35"

/dev\_stage="stage 25-35"

/lab\_host="E.coli DH10B"

/clone\_lib="Xenopus tropicalis xthr plasmid library"

/note="Vector: pCMVSPORT6\_xthr; Site\_1: SfiI; Site\_2:

SfiI; Xenopus tropicalis polyA+ RNA was obtained from pool

of heads and retinas from tailbud stages 25-35 cDNAs were

synthesized using the SMART system of CLONTECH and

directionally cloned into pCMVSPORT6\_xthr, a modified

version of pCMVSPORT6 allowing directional cloning using

asymetric SfiI sites. For antisense RNA synthesis, use T7

promoter and for sense RNA use SP6 promoter. Library

constructed by Drs. N. Pollet, M. Perron, M. Wegnez, A.

Mazabraud (CNRS UMR 8080, Universite Paris Sud, Orsay,

France)."

Query Match 87.0%; Score 17.4; DB 7; Length 638;

Best Local Similarity 94.7%; Pred. No. 2.5e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCGAAGAACAG 19

|||||

Db 402 CAAGTAATGCGAAGAACAG 384

|||||

RESULT 54

AZ600504

LOCUS

DEFINITION

1M0418C20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0418C20 F, genomic survey sequence.

AZ600504

ACCESSION

VERSION

KEYWORDS

GSS.

AZ600504.1 GI:11722694

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 638)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von

Niederhauser, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

# FEATURES

source

1. .638

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/strain="Ivory Coast"

/db\_xref="taxon:8364"

/clones="xthr34kl1"

/tissue\_type="pool of heads and retinas from tailbud

stages 25-35"

/dev\_stage="stage 25-35"

/lab\_host="E.coli DH10B"

/clone\_lib="Xenopus tropicalis xthr plasmid library"

/note="Vector: pCMVSPORT6\_xthr; Site\_1: SfiI; Site\_2:

SfiI; Xenopus tropicalis polyA+ RNA was obtained from pool

of heads and retinas from tailbud stages 25-35 cDNAs were

synthesized using the SMART system of CLONTECH and

directionally cloned into pCMVSPORT6\_xthr, a modified

version of pCMVSPORT6 allowing directional cloning using

asymetric SfiI sites. For antisense RNA synthesis, use T7

promoter and for sense RNA use SP6 promoter. Library

constructed by Drs. N. Pollet, M. Perron, M. Wegnez, A.

Mazabraud (CNRS UMR 8080, Universite Paris Sud, Orsay,

France)."

Query Match 87.0%; Score 17.4; DB 7; Length 638;

Best Local Similarity 94.7%; Pred. No. 2.5e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCGAAGAACAG 19

|||||

Db 402 CAAGTAATGCGAAGAACAG 384

|||||

RESULT 54

AZ600504

LOCUS

DEFINITION

1M0418C20F Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC1M0418C20 F, genomic survey sequence.

AZ600504

ACCESSION

VERSION

KEYWORDS

GSS.

AZ600504.1 GI:11722694

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;

Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 638)

Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,

Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,

Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von

Niederhauser, A. and Wright, D., Weiss, R.

Mouse whole genome scaffolding with paired end reads from 10kb

plasmid inserts

Unpublished (2000)

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah



Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA  
 Tel: 801 585 5606  
 Fax: 801 585 7177  
 Email: ddunn@genetics.utah.edu  
 Insert Length: 10000 Std Error: 0.00  
 Plate: 0418 row: C column: 20  
 Seq primer: CGTTGTAAACGACGCCAGT  
 Class: plasmid ends  
 High quality sequence stop: 638.  
 Location/Qualifiers  
 1. 638

#### FEATURES

source

/organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="UUGC1M0418C20"  
 /sex="Male"  
 /lab\_host="E. Coli strain XL10-Gold, TI-resistant, F-"  
 /clone\_lib="Mouse 10kb plasmid UUGC1M library"  
 /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (G1|4732114|GB|Af129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 9; Length 638;  
 Best Local Similarity 94.7%; Pred. No. 2.5e+03;  
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20

|||||  
 Db 280 AAGTAAATACAGAAACAGG 298

RESULT 55

LOCUS BX758905

DEFINITION BX758905 XGC-gastrula Xenopus tropicalis cDNA clone TGa131122 3', mRNA sequence.

ACCESSION BX758905

VERSION BX758905.1 GI:396666113

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 639)

REFERENCE Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

Sanger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

CONTACT: Croning MDR

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TGa131122.q1kT7

Sequencing primer: T7

This sequence is from a Xenopus Gene Collection (XGC) library constructed by Aaron M. Zorn.

cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli XL1-blue.

#### FEATURES

source

Location/Qualifiers

1. 639

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TGa131122"

/dev\_stage="gastrula (stages 10.5-12 mixed)"

/lab\_host="Escherichia coli XL1-blue"

/clone\_lib="XGC-gastrula"

/note="Vector: pCS107; Site 1: EcoRI; Site 2: NotI; cDNA was oligo dt primed from Sug of poly A+ RNA from stages 10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

#### ORIGIN

Query Match 87.0%; Score 17.4; DB 5; Length 639;

Best Local Similarity 94.7%; Pred. No. 2.5e+03;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19

|||||

Db 236 CAAGTAAATGAAGAAACAG 254

RESULT 56

LOCUS AL895824/c

DEFINITION AL895824 XGC-egg Xenopus tropicalis cDNA clone TEGg100k12 5', mRNA sequence.

ACCESSION AL895824

VERSION AL895824.2 GI:38695676

KEYWORDS EST.

SOURCE Xenopus tropicalis (western clawed frog)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae; Xenopus; Silurana.

1 (bases 1 to 649)

REFERENCE Croning,M.D.R., Ashurst,J.L., Taylor,R., Zorn,A.M. and Rogers,J.

Sanger Xenopus tropicalis EST project 2001 (11\_2003)

Unpublished (2003)

On Sep 16, 2002 this sequence version replaced gi:22948337.

CONTACT: Taylor R

Sanger Institute

Hinxton, Cambridgeshire, CB10 1SA, UK

Email: trop@sanger.ac.uk

Sanger Xenopus tropicalis EST project 2001

TROPICALIS\_SEQUENCE\_ID: TEGg100k12.plkSP6

Sequencing primer: Sp6

This sequence is from a Xenopus Gene Collection (XGC) library

constructed by Aaron M. Zorn.

cDNA was oligo dt primed from Sug of poly A+ RNA from egg.

EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end.

Vector: pCS107; Site 1: EcoRI; Site 2: NotI

Host: Escherichia coli XL1-blue.

#### FEATURES

source

Location/Qualifiers

1. 649

/organism="Xenopus tropicalis"

/mol\_type="mRNA"

/db\_xref="taxon:8364"

/clone="TEGg100k12"

/dev\_stage="egg"

/lab\_host="Escherichia coli XL1-blue"

```

/clone.lib="XGC-egg"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from egg.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end"

ORIGIN
Query Match      87.0%; Score 17.4; DB 1; Length 649;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCGAGAACAG 19
|||||
Db 406 CAAGTAATGAGAACAG 388

RESULT 57
CR426638
LOCUS
DEFINITION CR426638 XGC-tailbud Xenopus tropicalis cDNA clone TtBA070108 3',
mRNA sequence.
ACCESSION CR426638.1 GI:48920047
VERSION CR426638
KEYWORDS EST
SOURCE Xenopus tropicalis (western clawed frog)
ORGANISM Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 652)
Croning,M.D.R., Ashurst,J.L., Taylor,R., Garrett,N. and Rogers,J.
Sanger Xenopus tropicalis EST project 2001 (2004)
TROPICALIS_SEQUENCE ID: TtBA070108.g1kT7
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Nigel Garrett.
Seq primer: T7
Location/Qualifiers
1..652
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/db_xref="taxon:8364"
/clone="TtBA070108"
/dev_stage="tailbud (stage 28-30)"
/lab_host="Escherichia coli DH10B."
/clone_lib="XGC-tailbud"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from tailbud.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."

ORIGIN
Query Match      87.0%; Score 17.4; DB 7; Length 652;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCGAGAACAG 19
|||||
Db 237 CAAGTAATGAGAACAG 255

RESULT 58
CN084914
LOCUS
DEFINITION EC2BBA24DC01.g1 Xenopus tropicalis xtbs plasmid library Xenopus
tropicalis cDNA clone xtbs24F02 5', mRNA sequence.
ACCESSION CN084914
VERSION CN084914.1 GI:45877610

```

```

EST.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Silurana.
REFERENCE
1 (bases 1 to 658)
Thuret,R., Fierro,A.C., Coen,L., Perron,M., Demeneix,B., Wegnez,M.,
Gaspay,G., Weissenbach,J., Wincker,P., Mazabraud,A. and Pollet,N.
Exploring the nervous system transcriptome in the model Xenopus
tropicalis using EST analysis
Unpublished (2004)
Contact: Pollet N
Transgenese et Genetique des Amphibiens
CNRS UMR 8080
IBAIC bat 447, Universite Paris Sud, Orsay, F-91405, France
Tel: +33 169157272
Fax: +33 169156816
Email: Nicolas.Pollet@ibaic.u-psud.fr.
Location/Qualifiers
1..658
/organism="Xenopus tropicalis"
/mol_type="mRNA"
/strain="Ivory Coast"
/db_xref="taxon:8364"
/clone="xtbs24F02"
/tissue_type="pool of brains and spinal cords from
tadpoles at stages 51-52 and 61-62"
/dev_stage="stage 51-52 and 61-62"
/lab_host="E.coli DH10B"
/clone_lib="Xenopus tropicalis xtbs plasmid library"
/note="Vector: pCMVSPORT6_xtbs; Site_1: SfiI; Site_2:
SfiI; Xenopus tropicalis polyA+ RNA was obtained from
brain and spinal cord of tadpoles at stages 51-52 and
61-62. cDNAs were synthesized using the SMART system of
CLONTECH and directionally cloned into pCMVSPORT6 xtbs, a
modified version of pCMVSPORT6 allowing directional
cloning using asymmetric SfiI sites. For antisense RNA
synthesis, use T7 promoter and for sense RNA use SP6
promoter. Library constructed by Dr. L. Coen and Prof. B.
Demeneix (Museum National d'Histoire Naturelle and CNRS
UMR 5166, Paris, France)."

ORIGIN
Query Match      87.0%; Score 17.4; DB 7; Length 658;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CAAGTAATGCGAGAACAG 19
|||||
Db 434 CAAGTAATGAGAACAG 416

RESULT 59
BJ076021
LOCUS
DEFINITION BJ076021 NIBB Mochii normalized Xenopus tailbud library Xenopus
laevis cDNA clone XL104g09 3', mRNA sequence.
ACCESSION BJ076021
VERSION BJ076021.1 GI:17520937
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus; Xenopus.
REFERENCE
1 (bases 1 to 663)
Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and
Kohara,Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadaasu Shin-i
Center For Genetic Resource Information

```

National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshinigenes.nig.ac.jp  
The information of this clone is available through the following URL.  
<http://xenopus.nibb.ac.jp>.

Location/Qualifiers  
1..663  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XL104G09"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"  
/clone\_lib="NIBB Mochii normalized Xenopus tailbud library"

Query Match 87.0%; Score 17.4; DB 3; Length 663;  
Best Local Similarity 94.7%; Pred. NO. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
DB 197 CAAGTAAATGCAGAAACAG 215  
|||||

Query Match 87.0%; Score 17.4; DB 8; Length 669;  
Best Local Similarity 94.7%; Pred. NO. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
DB 232 CAAGTAAATGCAGAAACAG 250  
|||||

Search completed: January 28, 2006, 01:14:34  
Job time : 1628.9 secs

FEATURES  
source  
1..663  
/organism="Xenopus laevis"  
/mol\_type="mRNA"  
/db\_xref="taxon:8355"  
/clone="XL104G09"  
/tissue\_type="whole embryo"  
/dev\_stage="stage 25"  
/clone\_lib="NIBB Mochii normalized Xenopus tailbud library"

ORIGIN  
Query Match 87.0%; Score 17.4; DB 3; Length 663;  
Best Local Similarity 94.7%; Pred. NO. 2.5e+03;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
DB 197 CAAGTAAATGCAGAAACAG 215  
|||||

RESULT 60  
CX491676  
LOCUS  
DEFINITION  
IMAGE:7553198 3', mRNA sequence.  
CX491676  
CX491676.1 GI:57403338  
EST.  
Xenopus tropicalis (western clawed frog)  
Xenopus tropicalis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Xenopus; Silurana.  
1 (bases 1 to 669)  
Richardson, P., Lucas, S., Rokhsar, D., Dettter, J.C., Ng, D.C.,  
Brokstein, P. and Lindquist, E.A.  
DOE Joint Genome Institute Xenopus tropicalis EST project  
Unpublished (2004)  
Other ESTs: JGI XZG38345.fwd  
Contact: Lindquist, E.A., Richardson, P.  
DOE Joint Genome Institute  
2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
Tel: 925 296 5600  
Fax: 925 296 5710  
Email: [cdna@jgi-psf.org](mailto:cdna@jgi-psf.org)  
Tissue Procurement: Richard M. Harland Laboratory, University of  
California, Berkeley: <http://tropicalis.berkeley.edu/home>  
cDNA Library Preparation: Richard M. Harland Laboratory, University  
of California, Berkeley  
DNA Sequencing: DOE Joint Genome Institute: <http://www.jgi.doe.gov>  
Clone Distribution: I.M.A.G.E. Consortium/LLNL:  
<http://image.llnl.gov>  
Naming Conventions: EST name is generated by the concatenation of  
the JGI Clone id and the direction of sequencing. The suffix '.rev'  
indicates a reverse sequencing read of the insert. It does not  
necessarily reflect the orientation of the insert.  
Small Insert: Based upon one or more sequencing reads of this clone  
where vector sequence was present at both ends, this clone has been  
determined to contain a cDNA insert on the order of 600-1000 bases.  
Plate: XZG 0397 row: b column: 12  
High quality sequence stop: 669.  
Location/Qualifiers  
1..669  
/organism="Xenopus tropicalis"

FEATURES  
source  
1..669  
/organism="Xenopus tropicalis"

***This Page Blank (uspto)***

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 21:56:54 ; Search time 56.1798 Seconds  
(without alignments)  
632.812 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaaatgcagaaacagg 20

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 60 summaries

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
  - 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
  - 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
  - 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
  - 6: /cgn2\_6/ptodata/1/ina/PCUS COMB.seq:\*
  - 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
  - 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
  - 9: /cgn2\_6/ptodata/1/ina/backfileseq.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.4	87.0	160759	3	US-09-949-016-16514
2	16.8	84.0	2689	2	US-08-876-546A-15
3	16.8	84.0	2689	3	US-09-412-252-15
4	16.8	84.0	9633	9	US-09-412-252-15
5	16.8	84.0	9633	9	US-09-412-252-15
6	16.8	84.0	9726	3	US-09-869-588-22
7	16.8	84.0	24644	3	US-09-869-588-21
8	16.4	82.0	601	3	US-09-949-016-13113
9	16.4	82.0	601	3	US-09-949-016-10792
10	16.4	82.0	601	3	US-09-949-016-110815
11	16.4	82.0	601	3	US-09-949-016-110976
12	16.4	82.0	601	3	US-09-949-016-111137
13	16.4	82.0	714	3	US-09-949-016-111298
14	16.4	82.0	924	3	US-09-134-001C-2794
15	16.4	82.0	924	3	US-08-956-171E-745
16	16.4	82.0	1035	2	US-07-941-523-22
17	16.4	82.0	1110	2	US-07-941-523-20
18	16.4	82.0	1196	3	US-09-489-847-15
19	16.4	82.0	1638	2	US-08-838-219B-8
20	16.4	82.0	1638	3	US-09-233-336A-8
21	16.4	82.0	1638	3	US-09-233-336A-8
22	16.4	82.0	1638	3	US-09-402-036-8
23	16.4	82.0	1638	3	US-09-904-226-8
24	16.4	82.0	21438	3	US-09-949-016-14198

C 25	16.4	82.0	28473	3	US-08-961-527-83	Sequence 83, Appl
C 26	16.4	82.0	89268	3	US-09-949-016-13025	Sequence 12025, A
C 27	16.4	82.0	90544	3	US-09-949-016-13302	Sequence 13302, A
C 28	16.4	82.0	171700	3	US-09-949-016-12276	Sequence 12276, A
C 29	16.4	82.0	171701	3	US-09-949-016-15835	Sequence 15835, A
C 30	16.4	82.0	174170	3	US-09-949-016-14810	Sequence 14810, A
C 31	16.4	82.0	174170	3	US-09-949-016-14811	Sequence 14811, A
C 32	16.4	82.0	174259	3	US-09-949-016-11968	Sequence 11968, A
C 33	16.4	82.0	174262	3	US-09-949-016-14259	Sequence 14259, A
C 34	16.4	82.0	174318	3	US-09-949-016-11880	Sequence 11880, A
C 35	16.4	82.0	174318	3	US-09-949-016-14812	Sequence 14812, A
C 36	16.4	82.0	174318	3	US-09-949-016-14813	Sequence 14813, A
C 37	16.4	82.0	390890	3	US-09-949-016-14720	Sequence 14720, A
C 38	16	80.0	601	3	US-09-949-016-68433	Sequence 68433, A
C 39	16	80.0	601	3	US-09-949-016-68434	Sequence 68434, A
C 40	16	80.0	601	3	US-09-949-016-68435	Sequence 68435, A
C 41	16	80.0	601	3	US-09-949-016-68436	Sequence 68436, A
C 42	16	80.0	95566	3	US-09-949-016-11877	Sequence 11877, A
C 43	16	80.0	112112	3	US-09-949-016-15639	Sequence 15639, A
C 44	16	80.0	130724	3	US-09-949-016-13753	Sequence 13753, A
C 45	15.8	79.0	210	3	US-08-847-065-15	Sequence 15, Appl
C 46	15.8	79.0	210	3	US-09-829-382-15	Sequence 2385, Ap
C 47	15.8	79.0	510	3	US-09-583-110-2385	Sequence 274, App
C 48	15.8	79.0	561	3	US-09-107-433-274	Sequence 66214, A
C 49	15.8	79.0	601	3	US-09-949-016-66214	Sequence 107420, A
C 50	15.8	79.0	601	3	US-09-949-016-107420	Sequence 107489, A
C 51	15.8	79.0	601	3	US-09-949-016-140767	Sequence 140767, A
C 52	15.8	79.0	601	3	US-09-949-016-140768	Sequence 140768, A
C 53	15.8	79.0	601	3	US-09-949-002-2533	Sequence 2533, Ap
C 54	15.8	79.0	601	3	US-09-949-002-2534	Sequence 2534, Ap
C 55	15.8	79.0	601	3	US-09-949-002-2535	Sequence 2535, Ap
C 56	15.8	79.0	601	3	US-09-949-002-4046	Sequence 4046, Ap
C 57	15.8	79.0	601	3	US-09-949-002-4047	Sequence 4047, Ap
C 58	15.8	79.0	601	3	US-09-949-002-4048	Sequence 4048, Ap
C 59	15.8	79.0	601	3	US-09-949-002-4048	Sequence 2578, Ap
C 60	15.8	79.0	672	3	US-09-134-000C-2578	

ALIGNMENTS

RESULT 1  
US-09-949-016-16514  
; Sequence 16514, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16514  
; LENGTH: 160759  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(160759)  
; OTHER INFORMATION: n = A,T,C or G

US-09-949-016-16514  
Query Match 87.0%; Score 17.4; DB 3; Length 160759;  
Best Local Similarity 94.7%; Pred. No. 1.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
|||||  
Db 94050 CAAGTAAATTCAGAAACAG 94068

RESULT 2  
US-08-876-546A-15  
; Sequence 15, Application US/08876546A  
; Patent No. 5962665  
; GENERAL INFORMATION:  
; APPLICANT: P. Kroeger  
; APPLICANT: K. Abravaya  
; APPLICANT: J. Gorzowski  
; APPLICANT: R. Hoenle  
; APPLICANT: C. Esping  
; APPLICANT: J. Moore  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/876,546A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul D. Yaegeer  
; REGISTRATION NUMBER: 37,477  
; REFERENCE/DOCKET NUMBER: 6127.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/938-3508  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2689 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic RNA (HIV-2)  
US-08-876-546A-15

Query Match 84.0%; Score 16.8; DB 2; Length 2689;  
Best Local Similarity 90.0%; Pred. No. 2e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
Db 2139 CAAGTAAATGCAGAACTAGG 2158  
RESULT 3  
US-09-412-252-15  
; Sequence 15, Application US/09412252  
; Patent No. 6232455  
; GENERAL INFORMATION:  
; APPLICANT: P. Kroeger  
; APPLICANT: K. Abravaya  
; APPLICANT: J. Gorzowski  
; APPLICANT: R. Hoenle  
; APPLICANT: C. Esping  
; APPLICANT: J. Moore  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING

Query Match 84.0%; Score 16.8; DB 2; Length 2689;  
Best Local Similarity 90.0%; Pred. No. 2e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
Db 2139 CAAGTAAATGCAGAACTAGG 2158

RESULT 3  
US-09-412-252-15  
; Sequence 15, Application US/09412252  
; Patent No. 6232455  
; GENERAL INFORMATION:  
; APPLICANT: P. Kroeger  
; APPLICANT: K. Abravaya  
; APPLICANT: J. Gorzowski  
; APPLICANT: R. Hoenle  
; APPLICANT: C. Esping  
; APPLICANT: J. Moore  
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING

; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/412,252  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/876,546  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul D. Yaegeer  
; REGISTRATION NUMBER: 37,477  
; REFERENCE/DOCKET NUMBER: 6127.US.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/938-3508  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2689 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic RNA (HIV-2)  
US-09-412-252-15

Query Match 84.0%; Score 16.8; DB 3; Length 2689;  
Best Local Similarity 90.0%; Pred. No. 2e+02; Mismatches 0; Indels 0; Gaps 0;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
Db 2139 CAAGTAAATGCAGAACTAGG 2158

RESULT 4  
5223423-1  
; Patent No. 5223423  
; APPLICANT: FRANCHINI, GENOVEFFA;WONG-STAAAL, FLOSSIE;  
; GALLO, ROBERT  
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT  
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/331,212  
; FILING DATE: 03-31-1989  
; SEQ ID NO:1:  
; LENGTH: 9633  
5223423-1

Query Match 84.0%; Score 16.8; DB 9; Length 9633;  
Best Local Similarity 90.0%; Pred. No. 2.4e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 20  
|||||  
Db 4209 CAAGTAAATGCAGAACTAGG 4228

RESULT 5  
US-09-869-588-22  
; Sequence 22, Application US/09869588

```
; Patent No. 6790657
; GENERAL INFORMATION:
; APPLICANT: Atya
; TITLE OF INVENTION: Lentivirus Vector System
; FILE REFERENCE: 59316
; CURRENT APPLICATION NUMBER: US/09/869,588
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/US00/00390
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,247
; PRIOR FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; TYPE: DNA
; LENGTH: 9663
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: PCM-ROD(SD36/EM)
; NAME/KEY: misc_feature
; LOCATION: (1)..(9663)
; OTHER INFORMATION: n represents a, c, g or t.
US-09-869-588-22

Query Match      84.0%; Score 16.8; DB 3; Length 9663;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  CAAGTAAATGCAGAAACAGG 20
        |||||
Db      4818 CAAGTAAATGCAGAACTAGG 4837

RESULT 6
US-09-869-588-21
; Sequence 21, Application US/09869588
; Patent No. 6790657
; GENERAL INFORMATION:
; APPLICANT: Atya
; TITLE OF INVENTION: Lentivirus Vector System
; FILE REFERENCE: 59316
; CURRENT APPLICATION NUMBER: US/09/869,588
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/US00/00390
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,247
; PRIOR FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 9726
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PROD(SD36/EM)
; NAME/KEY: misc_feature
; LOCATION: (1)..(9726)
; OTHER INFORMATION: n represents a, c, t, or g.
US-09-869-588-21

Query Match      84.0%; Score 16.8; DB 3; Length 9726;
Best Local Similarity 90.0%; Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  CAAGTAAATGCAGAAACAGG 20
        |||||
Db      4214 CAAGTAAATGCAGAACTAGG 4233

RESULT 7
US-09-949-016-13113/c
; Sequence 13113, Application US/09949016
; Patent No. 6812339
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13113
; LENGTH: 246444
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)....(246444)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13113

Query Match      84.0%; Score 16.8; DB 3; Length 246444;
Best Local Similarity 90.0%; Pred. No. 3.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1  CAAGTAAATGCAGAAACAGG 20
        |||||
Db      36561 CAAGTAAAGGCAGAAATAGG 36542

RESULT 8
US-09-949-016-107692
; Sequence 107692, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107692
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
; NAME/KEY: misc_feature
; LOCATION: (1)....(601)
; OTHER INFORMATION: n represents a, c, t, or g.
US-09-949-016-107692

Query Match      82.0%; Score 16.4; DB 3; Length 601;
Best Local Similarity 94.4%; Pred. No. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2  AAGTAAATGCAGAAACAG 19
        |||||
Db      396 AAGAAATGCAGAAACAG 413

RESULT 9
US-09-949-016-110815/c
; Sequence 110815, Application US/09949016
; Patent No. 6812339
```

; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 110815  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-110815

Query Match 82.0%; Score 16.4; DB 3; Length 601;  
Best Local Similarity 94.4%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 192 AAGTAAATGCAGAAACAG 175

RESULT 10  
US-09-949-016-110976/c  
; Sequence 110976, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 110976  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-110976

Query Match 82.0%; Score 16.4; DB 3; Length 601;  
Best Local Similarity 94.4%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 192 AAGTAAATGCAGAAACAG 175

RESULT 11  
US-09-949-016-111137/c  
; Sequence 111137, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307

; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 111137  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-111137

Query Match 82.0%; Score 16.4; DB 3; Length 601;  
Best Local Similarity 94.4%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 192 AAGTAAATGCAGAAACAG 175

RESULT 12  
US-09-949-016-111298/c  
; Sequence 111298, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 111298  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-111298

Query Match 82.0%; Score 16.4; DB 3; Length 601;  
Best Local Similarity 94.4%; Pred. No. 2.6e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 192 AAGTAAATGCAGAAACAG 175

RESULT 13  
US-09-134-001C-2794/c  
; Sequence 2794, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779



```

; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2794
; LENGTH: 714
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2794

Query Match      82.0%; Score 16.4; DB 3; Length 714;
Best Local Similarity 94.4%; Pred. NO. 2.6e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACA 18
        |||||||
DB      137 CAAGTAAATGCAGAAATCA 120

RESULT 14
US-08-956-171E-745
; Sequence 745; Application US/08956171E
; Patent NO. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,171E
FILING DATE: 20-Oct-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,861
FILING DATE: January 5, 1996
APPLICATION NUMBER: 08/781,986
FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:
NAME: Mark J. Hyman
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB248P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (240) 314-1224
TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 745:
SEQUENCE CHARACTERISTICS:
LENGTH: 924 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 745:
US-08-956-171E-745

Query Match      82.0%; Score 16.4; DB 3; Length 924;
Best Local Similarity 94.4%; Pred. NO. 2.7e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACA 18
        |||||||

```

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/941,523
;   FILING DATE: 19920908
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Granahan, Patricia
;   REGISTRATION NUMBER: 32,227
;   REFERENCE/DOCKET NUMBER: BNL90-01A
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 861-6240
;   TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 22:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1035 base pairs
;     TYPE: NUCLEIC ACID
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
; US-07-941-523-22
;
; Query Match      82.0%; Score 16.4; DB 2; Length 1035;
; Best Local Similarity 94.4%; Pred. No. 2.7e+02;
; Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY      3 AGTAAATGCAGAAACAGG 20
;         ||||| ||||| |||||
; DB      246 AGTAAAGCAGAAACAGG 263
;
; RESULT 17
; US-07-941-523-40
; Sequence 20, Application US/07941523
; Patent No. 5571718
; GENERAL INFORMATION:
;   APPLICANT: Dunn, John J
;   APPLICANT: Barbour, Alan G
;   TITLE OF INVENTION: Cloning and Expression of Borrelia
;   TITLE OF INVENTION: Lipoproteins
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     CITY: Lexington
;     STATE: Massachusetts
;     COUNTRY: U.S.A.
;     ZIP: 01730
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/941,523
;   FILING DATE: 19920908
;   CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
;   NAME: Granahan, Patricia
;   REGISTRATION NUMBER: 32,227
;   REFERENCE/DOCKET NUMBER: BNL90-01A
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 861-6240
;   TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 20:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1110 base pairs
;     TYPE: NUCLEIC ACID
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
; US-07-941-523-20
;
; Query Match      82.0%; Score 16.4; DB 2; Length 1110;
; Best Local Similarity 94.4%; Pred. No. 2.8e+02;
; Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY      3 AGTAAATGCAGAAACAGG 20
;         ||||| ||||| |||||
; DB      321 AGTAAAGCAGAAACAGG 338
;
; RESULT 18
; US-09-489-847-15/c
; Sequence 15, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al
;   TITLE OF INVENTION: 98 Human Secreted Proteins
;   FILE REFERENCE: PZ031P1
;   CURRENT APPLICATION NUMBER: US/09/489,847
;   CURRENT FILING DATE: 2000-01-24
;   EARLIER APPLICATION NUMBER: PCT/US99/17130
;   EARLIER FILING DATE: 1999-07-29
;   EARLIER APPLICATION NUMBER: 60/094,657
;   EARLIER FILING DATE: 1998-07-30
;   EARLIER APPLICATION NUMBER: 60/095,486
;   EARLIER FILING DATE: 1998-08-05
;   EARLIER APPLICATION NUMBER: 60/096,319
;   EARLIER FILING DATE: 1998-08-12
;   EARLIER APPLICATION NUMBER: 60/095,454
;   EARLIER FILING DATE: 1998-08-06
;   EARLIER APPLICATION NUMBER: 60/095,455
;   EARLIER FILING DATE: 1998-08-06
;   NUMBER OF SEQ ID NOS: 376
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 15
;   TYPE: DNA
;   LENGTH: 1196
;   ORGANISM: Homo sapiens
; US-09-489-847-15
;
; Query Match      82.0%; Score 16.4; DB 3; Length 1196;
; Best Local Similarity 94.4%; Pred. No. 2.8e+02;
; Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY      1 CAAGTAAATGCAGAAACA 18
;         ||| ||||| |||||
; DB      311 CAAATAAATGCAGAAACA 294
;
; RESULT 19
; US-08-838-219B-8
; Sequence 8, Application US/08838219B
; Patent No. 5877012
; GENERAL INFORMATION:
;   APPLICANT: Warren, Gregory W
;   APPLICANT: Kozziel, Michael G
;   APPLICANT: Mullins, Martha A
;   APPLICANT: Nye, Gordon J
;   APPLICANT: Carr, Brian
;   APPLICANT: Desai, Nalini M
;   APPLICANT: Kostichka, N. Kristy
;   APPLICANT: Duck, Nicholas B
;   APPLICANT: Estruch, Juan J
;   TITLE OF INVENTION: A No. 5877012el Class of Proteins for the
;   TITLE OF INVENTION: Control of Plant Pests
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: CIBA-GEIGY Corporation
;     STREET: 7 Skyline Drive
;     CITY: Hawthorne
;     STATE: NY
;     COUNTRY: USA
;     ZIP: 10532
```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/838,219B  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/463,483  
FILING DATE: 06-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1638 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..1191  
OTHER INFORMATION: /product= "Translation of cDNA  
US-08-838-219B-8  
OTHER INFORMATION: encoding VIP3A(a) receptor"

Query Match 82.0%; Score 16.4; DB 2; Length 1638;  
Best Local Similarity 94.4%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 334 AAGTAGTGCAGAAACAG 351

RESULT 20  
US-09-233-336A-8  
Sequence 8, Application US/09233336A  
Patent No. 6107279  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: A No. 6107279el Class of Proteins for the  
CONTROL OF PLANT PESTS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne

STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/233,336A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/838,219  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1638 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..1191  
OTHER INFORMATION: /product= "Translation of cDNA  
US-09-233-336A-8  
OTHER INFORMATION: encoding VIP3A(a) receptor"

Query Match 82.0%; Score 16.4; DB 3; Length 1638;  
Best Local Similarity 94.4%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 334 AAGTAGTGCAGAAACAG 351

RESULT 21  
US-09-233-752A-8  
Sequence 8, Application US/09233752A  
Patent No. 6137033  
GENERAL INFORMATION:  
APPLICANT: Warren, Gregory W  
APPLICANT: Koziel, Michael G  
APPLICANT: Mullins, Martha A  
APPLICANT: Nye, Gordon J  
APPLICANT: Carr, Brian  
APPLICANT: Desai, Nalini M  
APPLICANT: Kostichka, N. Kristy  
APPLICANT: Duck, Nicholas B  
APPLICANT: Estruch, Juan J  
TITLE OF INVENTION: A No. 6137033el Class of Proteins for the  
CONTROL OF PLANT PESTS  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:

ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/233,752A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/838,219  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,594  
FILING DATE: 09-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,018  
FILING DATE: 23-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/037,057  
FILING DATE: 25-MAR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Pace, Gary M.  
REGISTRATION NUMBER: 40,403  
REFERENCE/DOCKET NUMBER: CGC 1925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8582  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1638 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2..1191  
OTHER INFORMATION: /product= "Translation of cDNA encoding VIP3A(a) receptor"  
US-09-233-752A-8

Query Match 82.0%; Score 16.4; DB 3; Length 1638;  
Best Local Similarity 94.4%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
||||| |||||||  
Db 334 AAGTAGATGCAGAAACAG 351

RESULT 22  
US-09-402-036-8  
Sequence 8, Application US/09402036  
Patent No. 6291156  
GENERAL INFORMATION:  
APPLICANT: Estruch, Juan J.  
APPLICANT: Yu, Cao-Guo  
APPLICANT: Warren, Gregory W.  
APPLICANT: Desai, Nalini  
APPLICANT: Kozziel, Michael  
APPLICANT: Nye, Gordon  
TITLE OF INVENTION: Plant Pest Control  
FILE REFERENCE: S-21284C  
CURRENT APPLICATION NUMBER: US/09/402,036  
CURRENT FILING DATE: 2000-02-08

PRIOR APPLICATION NUMBER: PCT/EP98/01952  
PRIOR FILING DATE: 1998-04-02  
PRIOR APPLICATION NUMBER: 08/838,219  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,263  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,265  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/463,483  
PRIOR FILING DATE: 1995-06-05  
PRIOR APPLICATION NUMBER: 08/314,594  
PRIOR FILING DATE: 1994-09-09  
PRIOR APPLICATION NUMBER: 08/218,018  
PRIOR FILING DATE: 1994-03-24  
PRIOR APPLICATION NUMBER: 08/037,057  
PRIOR FILING DATE: 1993-03-25  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 1638  
TYPE: DNA  
ORGANISM: Agrotis ipsilon  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2)..(1189)  
OTHER INFORMATION: Translation of cDNA encoding VIP3A(a) receptor  
US-09-402-036-8

Query Match 82.0%; Score 16.4; DB 3; Length 1638;  
Best Local Similarity 94.4%; Pred. No. 2.9e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
||||| |||||||  
Db 334 AAGTAGATGCAGAAACAG 351

RESULT 23  
US-09-904-226-8  
Sequence 8, Application US/09904226  
Patent No. 6429360  
GENERAL INFORMATION:  
APPLICANT: Estruch, Juan J.  
APPLICANT: Warren, Gregory W.  
APPLICANT: Desai, Nalini  
APPLICANT: Kozziel, Michael  
APPLICANT: Nye, Gordon  
TITLE OF INVENTION: Plant Pest Control  
FILE REFERENCE: S-21284D  
CURRENT APPLICATION NUMBER: US/09/904,226  
CURRENT FILING DATE: 2001-07-12  
PRIOR APPLICATION NUMBER: PCT/EP98/01952  
PRIOR FILING DATE: 1998-04-02  
PRIOR APPLICATION NUMBER: 08/838,219  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,263  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/832,265  
PRIOR FILING DATE: 1997-04-03  
PRIOR APPLICATION NUMBER: 08/463,483  
PRIOR FILING DATE: 1995-06-05  
PRIOR APPLICATION NUMBER: 08/314,594  
PRIOR FILING DATE: 1994-09-09  
PRIOR APPLICATION NUMBER: 08/218,018  
PRIOR FILING DATE: 1994-03-24  
PRIOR APPLICATION NUMBER: 08/037,057  
PRIOR FILING DATE: 1993-03-25  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 1638  
TYPE: DNA

```
; ORGANISM: Agrotis ipsilon
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(1189)
; OTHER INFORMATION: Translation of cDNA encoding VIP3A(a) receptor
; OTHER INFORMATION: from Black cutworm
US-09-904-226-8

Query Match      82.0%; Score 16.4; DB 3; Length 1638;
Best Local Similarity 94.4%; Pred. No. 2.9e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      334 AAGTAGATGCAGAAACAG 351

RESULT 24
US-09-949-016-14198/c
; Sequence 14198, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14198
; LENGTH: 21438
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14198

Query Match      82.0%; Score 16.4; DB 3; Length 21438;
Best Local Similarity 94.4%; Pred. No. 4.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      14869 AAGTAAATGAAGAAACAG 14852

RESULT 25
US-08-961-527-83/c
; Sequence 83, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
```

```
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 83:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28473 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-83

Query Match      82.0%; Score 16.4; DB 3; Length 28473;
Best Local Similarity 94.4%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      16919 AAGTAAATGCAGTAACAG 16902

RESULT 26
US-09-949-016-12025/c
; Sequence 12025, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12025
; LENGTH: 89268
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12025

Query Match      82.0%; Score 16.4; DB 3; Length 89268;
Best Local Similarity 94.4%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
      ||||| ||||| ||||| |||||
Db      33178 AAGTAAAGGCAGAAACAG 33161

RESULT 27
US-09-949-016-13302/c
; Sequence 13302, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/949,016
```

; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13302  
; LENGTH: 90544  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-13302

Query Match 82.0%; Score 16.4; DB 3; Length 90544;  
Best Local Similarity 94.4%; Pred. No. 4.8e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
||||| |||||||  
DB 33363 AAGTAAAGCGAGAAACAG 33346

RESULT 28  
US-09-949-016-12276/c  
; Sequence 12276, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12276  
; LENGTH: 171700  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12276

Query Match 82.0%; Score 16.4; DB 3; Length 171700;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 18  
||||| |||||||  
DB 45460 CAAGTAAATGCATAAACA 45443

RESULT 29  
US-09-949-016-15835/c  
; Sequence 15835, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03

; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15835  
; LENGTH: 171701  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15835

Query Match 82.0%; Score 16.4; DB 3; Length 171701;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 18  
||||| |||||||  
DB 45460 CAAGTAAATGCATAAACA 45443

RESULT 30  
US-09-949-016-14810/c  
; Sequence 14810, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14810  
; LENGTH: 174170  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(174170)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14810

Query Match 82.0%; Score 16.4; DB 3; Length 174170;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
||||| |||||||  
DB 94124 AAGTAAATGCAGAAACAG 94107

RESULT 31  
US-09-949-016-14811/c  
; Sequence 14811, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498

; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14811  
; LENGTH: 174170  
; TYPE: DNA  
; ORGANISM: Human  
; NAME/KEY: misc feature  
; LOCATION: (1)-(174170)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-14811

Query Match 82.0%; Score 16.4; DB 3; Length 174170;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 94124 AAGTAAATGCAGAAACAG 94107

## RESULT 32

US-09-949-016-11968  
; Sequence 11968, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11968  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11968

Query Match 82.0%; Score 16.4; DB 3; Length 174259;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 105890 AAGTAAATGCAGAAACAG 105907

## RESULT 33

US-09-949-016-14259  
; Sequence 14259, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08

; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14259  
; LENGTH: 174262  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-14259

Query Match 82.0%; Score 16.4; DB 3; Length 174262;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 105890 AAGTAAATGCAGAAACAG 105907

## RESULT 34

US-09-949-016-11880/c  
; Sequence 11880, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11880  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)-(174318)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-11880

Query Match 82.0%; Score 16.4; DB 3; Length 174318;  
Best Local Similarity 94.4%; Pred. No. 5.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
DB 94272 AAGTAAATGCAGAAACAG 94255

## RESULT 35

US-09-949-016-14812/c  
; Sequence 14812, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012





;  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68434  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-68434

Query Match  
Best Local Similarity 80.0%; Score 16; DB 3; Length 601;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAATGCAGAAACAGG 20  
|||||

Db 239 TAAATGCAGAAACAGG 224  
|||||

RESULT 40  
US-09-949-016-68435/c  
; Sequence 68435, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68435  
; LENGTH: 601  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-68435

Query Match  
Best Local Similarity 80.0%; Score 16; DB 3; Length 601;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAATGCAGAAACAGG 20  
|||||

Db 465 TAAATGCAGAAACAGG 450  
|||||

RESULT 41  
US-09-949-016-68436/c  
; Sequence 68436, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 68436  
; LENGTH: 601

;  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-68436

Query Match  
Best Local Similarity 80.0%; Score 16; DB 3; Length 601;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAATGCAGAAACAGG 20  
|||||

Db 486 TAAATGCAGAAACAGG 471  
|||||

RESULT 42  
US-09-949-016-11877  
; Sequence 11877, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11877  
; LENGTH: 95566  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-11877

Query Match  
Best Local Similarity 80.0%; Score 16; DB 3; Length 95566;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GTAAATGCAGAAACAG 19  
|||||

Db 14408 GTAAATGCAGAAACAG 14423  
|||||

RESULT 43  
US-09-949-016-15639/c  
; Sequence 15639, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15639  
; LENGTH: 112112  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-15639

Query Match  
80.0%; Score 16; DB 3; Length 112112;

```
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACA 18
Db 5625 AGTAAATGCAGAAACA 5610

RESULT 44
US-09-949-016-13753
; Sequence 13753, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13753
; LENGTH: 130724
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(130724)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-13753

Query Match 80.0%; Score 16; DB 3; Length 130724;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 TAAATGCAGAAACAGG 20
Db 8750 TAAATGCAGAAACAGG 8765

RESULT 45
US-08-847-065-15
; Sequence 15, Application US/08847065
; Patent No. 6245335
; GENERAL INFORMATION:
; APPLICANT: Masure, H. Robert
; APPLICANT: Rosenow, Carsten I.
; APPLICANT: Tuomanen, Elaine
; APPLICANT: Wizemann, Theresa M.
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSES: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; FLOOR
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 08/847,065
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-158 ..
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 base pairs
```

```
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-158 ..
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: B. subtilis
US-08-847-065-15

Query Match 79.0%; Score 15.8; DB 3; Length 210;
Best Local Similarity 89.5%; Pred. No. 4.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20
Db 29 AAGTATATACAGAAACAGG 47

RESULT 46
US-09-829-382-15
; Sequence 15, Application US/09829382
; Patent No. 6784164
; GENERAL INFORMATION:
; APPLICANT: Masure, H. Robert
; APPLICANT: Rosenow, Carsten I.
; APPLICANT: Tuomanen, Elaine
; APPLICANT: Wizemann, Theresa M.
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSES: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; FLOOR
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/829,382
; FILING DATE: 09-Apr-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/847,065
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-158 ..
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 base pairs
```

```

;
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 274:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1..561
; SEQUENCE DESCRIPTION: SEQ ID NO: 274:
;
US-09-107-433-274

Query Match 79.0%; Score 15.8; DB 3; Length 561;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
||| ||||| |||||
DB 553 CAATTAATGCAGTAACAG 535

RESULT 49
US-09-949-016-66214
; Sequence 66214, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66214
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
;
US-09-949-016-66214

Query Match 79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19

```

```
Db      485 CAAAGTAAATGCAGGAACAG 503
||||| ||||| ||||| |||||
US-09-949-016-107420
; Sequence 107420, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107420
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-107420

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAAGTAATGCAGAAACAG 19
||||| ||||| ||||| |||||
Db      308 CAAATTAATGCAGAAACAG 326

US-09-949-016-107489
; Sequence 107489, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107489
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-107489

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAAGTAATGCAGAAACAG 19
||||| ||||| ||||| |||||
Db      308 CAAATTAATGCAGAAACAG 326

US-09-949-016-107499
; Sequence 107499, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107499
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-107499

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAAGTAATGCAGAAACAG 19
||||| ||||| ||||| |||||
Db      308 CAAATTAATGCAGAAACAG 326

US-09-949-016-107499
; Sequence 107499, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 107499
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-107499

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAAGTAATGCAGAAACAG 19
||||| ||||| ||||| |||||
Db      308 CAAATTAATGCAGAAACAG 326

US-09-949-016-140767
; Sequence 140767, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140767
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-140767

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAG 20
||||| ||||| ||||| |||||
Db      20 AAGTAACTTCAGAAACAG 38

US-09-949-016-140768
; Sequence 140768, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140768
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-140768

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAG 20
||||| ||||| ||||| |||||
Db      175 AAGTAACTTCAGAAACAG 193

US-09-949-002-2533/c
; Sequence 2533, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
```

```
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2533
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-2533

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 465 CAAATTAATGCAGAAACAG 447

RESULT 55
US-09-949-002-2534/c
; Sequence 2534, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2534
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-2534

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 465 CAAATTAATGCAGAAACAG 447

RESULT 56
US-09-949-002-2535/c
; Sequence 2535, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2535
```

```
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-2535

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 82 CAAATTAATGCAGAAACAG 64

RESULT 57
US-09-949-002-4046/c
; Sequence 4046, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4046
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-4046

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 465 CAAATTAATGCAGAAACAG 447

RESULT 58
US-09-949-002-4047/c
; Sequence 4047, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4047
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-4047

Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 465 CAAATTAATGCAGAAACAG 447
```

Db 143 CAAATTATGCAGAAACAG 125

## RESULT 59

```

US-09-949-002-4048/c
; Sequence 4048, Application US/09949002
; Patent No. 6900016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH INFLAMMATORY AUTOIMMUNE DISEASE, METHODS OF DETECTION
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000790
; CURRENT APPLICATION NUMBER: US/09/949,002
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/231,401
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 10823
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 4048
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-002-4048

```

```
Query Match          79.0%; Score 15.8; DB 3; Length 601;
Best Local Similarity 89.5%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

Qy 1 CAAGTAAATGCAGAAACAG 19  
Db 82 CAAATTAAATGCAGAAACAG 64

## RESULT 60

```

US-09-134-000C-2578
; Sequence 2578, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134, 000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2578
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2578

```

Query Match 79.0%; Score 15.8; DB 3; Length 672;  
Best Local Similarity 89.5%; Pred. No. 4.9e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAGG 20  
||| ||| ||| ||| |||  
Db 374 AAGTGAACGCAGAAACAGG 392

Search completed: January 28, 2006, 01:18:50  
Job time : 62.1798 secs



```
RESULT 2
US-09-815-242-144/c
; Sequence 144, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 130
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-10-282-122A-130
Query Match      87.0%; Score 17.4; DB 3; Length 272;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCGAAGAACAG 19
      |||||
DB      166 CAAGTAAATGAGAGAACAG 148
      |||||

RESULT 3
US-10-282-122A-130/c
; Sequence 130, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

Query Match      87.0%; Score 17.4; DB 3; Length 272;
Best Local Similarity 94.7%; Pred. No. 5.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCGAAGAACAG 19
      |||||
DB      166 CAAGTAAATGAGAGAACAG 148
      |||||

RESULT 4
US-10-027-632-193671
; Sequence 193671, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193671
; LENGTH: 442
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-193671
Query Match      87.0%; Score 17.4; DB 5; Length 442;
Best Local Similarity 94.7%; Pred. No. 5.7e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```



QY 2 AAGTAAATGCAGAAACAGG 20  
||||||| |||||||  
DB 247 AAGTAAATACAGAAACAGG 265

## RESULT 5

US-10-027-632-193671  
; Sequence 193671, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 193671  
; LENGTH: 442  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-193671

Query Match 87.0%; Score 17.4; DB 6; Length 442;  
Best Local Similarity 94.7%; Pred. No. 5.7e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
||||||| |||||||  
DB 247 AAGTAAATACAGAAACAGG 265

## RESULT 6

US-09-925-065A-767188/c  
; Sequence 767188, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 767188  
; LENGTH: 543  
; TYPE: DNA

; ORGANISM: Homo sapiens  
US-09-925-065A-767188

Query Match 87.0%; Score 17.4; DB 4; Length 543;  
Best Local Similarity 94.7%; Pred. No. 5.8e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 20  
||||||| |||||||  
DB 530 AAGAAATGCAGAAACAGG 512

## RESULT 7

US-10-282-122A-6267  
; Sequence 6267, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangau  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6267  
; LENGTH: 732  
; TYPE: DNA  
; ORGANISM: Enterococcus faecalis  
US-10-282-122A-6267

Query Match 87.0%; Score 17.4; DB 7; Length 732;  
Best Local Similarity 94.7%; Pred. No. 6.1e+02;  
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
||||||| |||||||  
DB 118 CAAGTAAATGCAGAAACAG 136

## RESULT 8

US-11-097-143-13502/c

```
; Sequence 13502, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13502
; LENGTH: 4911
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-13502

Query Match      87.0%; Score 17.4; DB 10; Length 4911;
Best Local Similarity 94.7%; Pred. No. 8.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAGG 20
      |||||
Db      1405 AGGTAATGCAGAAACAGG 1387

RESULT 9
US-11-097-143-13501/c
; Sequence 13501, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; FILE REFERENCE: DROSOPHILA GENES.
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
```

```
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13501
; LENGTH: 7886
; TYPE: DNA
; ORGANISM: DROSOPHILA
US-11-097-143-13501

Query Match      87.0%; Score 17.4; DB 10; Length 7886;
Best Local Similarity 94.7%; Pred. No. 9e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAGG 20
      |||||
Db      3324 AGGTAATGCAGAAACAGG 3306

RESULT 10
US-10-656-394A-13
; Sequence 13, Application US/10656394A
; Publication No. US20040210957A1
; GENERAL INFORMATION:
; APPLICANT: Wang et al.
; TITLE OF INVENTION: Cloning and Characterization of the
; FILE REFERENCE: broad-spectrum resistance gene Pi2
; CURRENT APPLICATION NUMBER: US/10/656,394A
; CURRENT FILING DATE: 2003-09-05
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 99090
; TYPE: DNA
; ORGANISM: Oryza minuta
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 23216
; OTHER INFORMATION: n = A,T,C or G
US-10-656-394A-13

Query Match      87.0%; Score 17.4; DB 8; Length 99090;
Best Local Similarity 94.7%; Pred. No. 1.4e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAGG 20
      |||||
Db      32637 AAGTAAATGCAGAGCAGG 32655

RESULT 11
US-10-087-192-1438
; Sequence 1438, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1438
; LENGTH: 177587
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
```

```
; LOCATION: (1)...(177587)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-1438

Query Match      87.0%; Score 17.4; DB 5; Length 177587;
Best Local Similarity 94.7%; Pred. No. 1.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   |||||
Db 102050 CAAGTAAATGCAGAAACAG 102068

RESULT 12
US-09-925-065A-332186
; Sequence 332186, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332186
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-332186

Query Match      85.0%; Score 17; DB 4; Length 603;
Best Local Similarity 100.0%; Pred. No. 9.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19
   |||||
Db 172 AGTAAATGCAGAAACAG 188

RESULT 13
US-09-925-065A-332187
; Sequence 332187, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: Nucleotide Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332187
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-332187

Query Match      85.0%; Score 17; DB 5; Length 703;
Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 18
   |||||
Db 196 AAGTAAATGCAGAAACAG 180

RESULT 15
US-10-027-632-127740/c
; Sequence 127740, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
```

```
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127740
; LENGTH: 703
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-127740

Query Match      85.0%; Score 17; DB 6; Length 703;
Best Local Similarity 100.0%; Pred. No. 9.3e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2  AAGTAAATGCAGAAACA 18
Db      196  AAGTAAATGCAGAAACA 180

RESULT 16
US-09-070-927A-773/c
; Sequence 773, Application US/09070927A
; Patent No. US20020120118A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
;          Patrick J. Dillon
;          Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 773:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1081 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
; SEQUENCE DESCRIPTION: SEQ ID NO: 773:
US-09-070-927A-773

Query Match      85.0%; Score 17; DB 3; Length 1081;
Best Local Similarity 89.5%; Pred. No. 1e+03;
Matches 17; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1  CAAGTAAATGCAGAAACAG 19
Db      608  CAAGTAAATGARGAAACAG 590

RESULT 17
US-09-925-065A-97556
; Sequence 97556, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97556
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-97556

Query Match      84.0%; Score 16.8; DB 4; Length 426;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1  CAAGTAAATGCAGAAACAG 20
Db      246  CAAGTAAATGCAGAAACAG 265

RESULT 18
US-09-925-065A-136681
; Sequence 136681, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
```



us-10-716-005-4.rnpbm

Mon Jan 30 07:42:55 2006

```
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429130
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-429130

Query Match      84.0%; Score 16.8; DB 4; Length 583;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAGG 20
      ||||| ||||| ||||| |||||
Db      398 CAAGAAATGCAGAAACAGG 417

RESULT 23
US-09-925-065A-429131
; Sequence 429131, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429131
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-429131

Query Match      84.0%; Score 16.8; DB 4; Length 583;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAGG 20
      ||||| ||||| ||||| |||||
Db      398 CAAGAAATGCAGAAACAGG 417

RESULT 24
US-09-925-065A-259716
; Sequence 259716, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 429131
; LENGTH: 583
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-45434
; Sequence 45434, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE OF INVENTION: Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 45434
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-45434

Query Match      84.0%; Score 16.8; DB 4; Length 614;
Best Local Similarity 90.0%; Pred. No. 1.1e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 CAAGTAAATGCAGAAACAGG 20
      ||||| ||||| ||||| |||||
Db      327 CAAGTAAATGCAGACACAGG 346

RESULT 26
US-10-027-632-289556
; Sequence 289556, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
```

; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 289556  
 ; LENGTH: 615  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-289556

Query Match 84.0%; Score 16.8; DB 5; Length 615;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
 || |||||  
 DB 442 CATGTAATGCAGAAACATG 461

RESULT 27  
 US-10-027-632-289556  
 ; Sequence 289556, Application US/10027632  
 ; Publication No. US20030204075A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 289556  
 ; LENGTH: 615  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-289556

Query Match 84.0%; Score 16.8; DB 6; Length 615;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
 || |||||  
 DB 442 CATGTAATGCAGAAACATG 461

RESULT 28  
 US-09-925-065A-853796/c  
 ; Sequence 853796, Application US/09925065A  
 ; Publication No. US20050228172A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single  
 ; FILE REFERENCE: 108827.135  
 ; CURRENT APPLICATION NUMBER: US/09/925,065A  
 ; CURRENT FILING DATE: 2001-08-08  
 ; PRIOR APPLICATION NUMBER: US 60/243,096  
 ; PRIOR FILING DATE: 2000-10-24  
 ; PRIOR APPLICATION NUMBER: US 60/252,147  
 ; PRIOR FILING DATE: 2000-11-20  
 ; PRIOR APPLICATION NUMBER: US 60/250,092  
 ; PRIOR FILING DATE: 2000-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/261,766  
 ; PRIOR FILING DATE: 2001-01-16  
 ; PRIOR APPLICATION NUMBER: US 60/289,846  
 ; PRIOR FILING DATE: 2001-05-09  
 ; NUMBER OF SEQ ID NOS: 957086  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 853796  
 ; LENGTH: 658  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-925-065A-853796

Query Match 84.0%; Score 16.8; DB 4; Length 658;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
 || |||||  
 DB 357 CAAGTTAATGCAGAAACAGG 338

RESULT 29  
 US-10-027-632-101447  
 ; Sequence 101447, Application US/10027632  
 ; Publication No. US20020198371A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, David G.  
 ; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 ; FILE REFERENCE: 108827.129  
 ; CURRENT APPLICATION NUMBER: US/10/027,632  
 ; CURRENT FILING DATE: 2002-04-30  
 ; PRIOR APPLICATION NUMBER: US 60/218,006  
 ; PRIOR FILING DATE: 2000-07-12  
 ; PRIOR APPLICATION NUMBER: US 60/198,676  
 ; PRIOR FILING DATE: 2000-04-20  
 ; PRIOR APPLICATION NUMBER: US 60/193,483  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: US 60/185,218  
 ; PRIOR FILING DATE: 2000-02-24  
 ; PRIOR APPLICATION NUMBER: US 60/167,363  
 ; PRIOR FILING DATE: 1999-11-23  
 ; PRIOR APPLICATION NUMBER: US 60/156,358  
 ; PRIOR FILING DATE: 1999-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/146,002  
 ; PRIOR FILING DATE: 1999-08-09  
 ; NUMBER OF SEQ ID NOS: 325720  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 101447  
 ; LENGTH: 692  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 US-10-027-632-101447

Query Match 84.0%; Score 16.8; DB 5; Length 692;  
 Best Local Similarity 90.0%; Pred. No. 1.1e+03;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
|||||  
Db 248 CAAGAAATGCAGATCAGG 267

RESULT 30  
US-10-027-632-101447  
; Sequence 101447, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 101447  
; LENGTH: 692  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-101447

Query Match 84.0%; Score 16.8; DB 6; Length 692;  
Best Local Similarity 90.0%; Pred. No. 1.1e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
|||||  
Db 248 CAAGAAATGCAGATCAGG 267

RESULT 31  
US-10-425-115-161074/c  
; Sequence 161074, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 161074  
; LENGTH: 821  
; TYPE: DNA  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(821)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:

; OTHER INFORMATION: Clone ID: MRT4577\_78481C.1  
US-10-425-115-161074

Query Match 84.0%; Score 16.8; DB 8; Length 821;  
Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
|||||  
Db 627 CAAGCAAGTCAGAAACAGG 608

RESULT 32  
US-09-925-065A-550161  
; Sequence 550161, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 550161  
; LENGTH: 843  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-925-065A-550161

Query Match 84.0%; Score 16.8; DB 4; Length 843;  
Best Local Similarity 90.0%; Pred. No. 1.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
|||||  
Db 721 CAAGTGAATGCAAAACAGG 740

RESULT 33  
US-09-925-065A-550162  
; Sequence 550162, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135  
; CURRENT APPLICATION NUMBER: US/09/925,065A  
; CURRENT FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: US 60/243,096  
; PRIOR FILING DATE: 2000-10-24  
; PRIOR APPLICATION NUMBER: US 60/252,147  
; PRIOR FILING DATE: 2000-11-20  
; PRIOR APPLICATION NUMBER: US 60/250,092  
; PRIOR FILING DATE: 2000-11-30  
; PRIOR APPLICATION NUMBER: US 60/261,766  
; PRIOR FILING DATE: 2001-01-16  
; PRIOR APPLICATION NUMBER: US 60/289,846  
; PRIOR FILING DATE: 2001-05-09  
; NUMBER OF SEQ ID NOS: 957086  
; SOFTWARE: FastSeq for Windows Version 4.0



```
; SEQ ID NO 550162
; LENGTH: 843
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-925-065A-550162

Query Match      84.0%; Score 16.8; DB 4; Length 843;
Best Local Similarity 90.0%; Pred. No. 1.2e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 721 CAAGTGAATGCAAAAACAGG 740

RESULT 34
US-10-437-963-26351
; Sequence 26351, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 26351
; LENGTH: 1812
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3114C.1
US-10-437-963-26351

Query Match      84.0%; Score 16.8; DB 7; Length 1812;
Best Local Similarity 90.0%; Pred. No. 1.3e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 451 CAAGCAATGCAGCAACAGG 470

RESULT 35
US-10-731-988-22
; Sequence 22, Application US/10731988
; Publication No. US20040147026A1
; GENERAL INFORMATION:
; APPLICANT: Arya
; TITLE OF INVENTION: Lentivirus Vector System
; FILE REFERENCE: 67517
; CURRENT APPLICATION NUMBER: US/10/731,988
; CURRENT FILING DATE: 2003-12-09
; PRIOR APPLICATION NUMBER: 09/869,588
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: PCT/US00/00390
; PRIOR FILING DATE: 2000-01-06
; PRIOR APPLICATION NUMBER: 60/115,247
; PRIOR FILING DATE: 1999-01-07
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 9663
; TYPE: DNA
; ORGANISM: Artificial Sequence

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: PROD(SD36/EM)
US-10-731-988-21

Query Match      84.0%; Score 16.8; DB 7; Length 9726;
Best Local Similarity 90.0%; Pred. No. 1.8e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
    ||||| ||||| ||||| |||||
Db 4214 CAAGTAAATGCAGAACTAGG 4233

RESULT 37
US-10-017-161-1989/c
; Sequence 1989, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABEURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1989
; LENGTH: 14241
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(14241)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(267)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4723)..(4988)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5026)..(5670)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (113871)..(14041)
US-10-017-161-1989

Query Match      84.0%; Score 16.8; DB 6; Length 14241;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 4047 CAAGTAAATGCAGAAACAGG 4028

RESULT 38
US-10-292-798-1637/c
; Sequence 1637, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1637
; LENGTH: 14425
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(14425)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(235)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (306)..(436)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4906)..(5171)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5209)..(5853)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (14055)..(14225)

US-10-292-798-1637
Query Match      84.0%; Score 16.8; DB 6; Length 14425;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 4230 CAAGTAAATGCAGAAACAGG 4211

RESULT 39
US-10-017-161-2251/c
; Sequence 2251, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2251
; LENGTH: 16270
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(16270)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(286)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2151)..(2281)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (6751)..(7016)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7054)..(7698)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (15900)..(16070)
US-10-017-161-2251

Query Match      84.0%; Score 16.8; DB 6; Length 16270;
Best Local Similarity 90.0%; Pred. No. 1.9e+03;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   |||||
Db 6075 CAAGTAAATGCAGAAACAGG 6056

RESULT 40
US-10-292-798-1897/c
; Sequence 1897, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
```

RESULT 42  
US-10-719-993-6799/c  
; Sequence 6799, Application US/10719993  
; Publication No. US20040265849A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001496  
; CURRENT APPLICATION NUMBER: US/10/719,993  
; CURRENT FILING DATE: 2003-11-24  
; NUMBER OF SEQ ID NOS: 55342  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6799  
; LENGTH: 255439  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-719-993-6799

Query Match 84.0%; Score 16.8; DB 8; Length 255439;  
Best Local Similarity 90.0%; Pred. No. 2.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
DB 40511 CAAGTAAAGGCAGAAATAGG 40492

RESULT 43  
US-10-424-599-27588/c  
; Sequence 27588, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 27588  
; LENGTH: 275  
; TYPE: DNA  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(275)  
; OTHER INFORMATION: unsure at all n locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_124913C.1  
US-10-424-599-27588

Query Match 82.0%; Score 16.4; DB 7; Length 275;  
Best Local Similarity 94.4%; Pred. No. 1.5e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAGG 19  
||||| ||||| ||||| ||||| |||||  
DB 29 AATTAAATGCAGAAACAGG 12

RESULT 44  
US-09-925-065A-168362  
; Sequence 168362, Application US/09925065A  
; Publication No. US20050228172A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single  
; TITLE OF INVENTION: Nucleotide Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.135

US-10-716-005-4.rnpbm  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 10/017,161  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: JP 2001-246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2070  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1897  
; LENGTH: 16270  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; LOCATION: source  
; FEATURE:  
; LOCATION: (1)..(16270)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (201)..(286)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2151)..(2281)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (6751)..(7016)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (7054)..(7698)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (15900)..(16070)  
US-10-292-798-1897

Query Match 84.0%; Score 16.8; DB 6; Length 16270;  
Best Local Similarity 90.0%; Pred. No. 1.9e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
DB 6075 CAAGTAAATGCAGAAACAGG 6056

RESULT 41  
US-10-322-281-516  
; Sequence 516, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 516  
; LENGTH: 41079  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(41079)  
; OTHER INFORMATION: n = A, T, C or G  
US-10-322-281-516

Query Match 84.0%; Score 16.8; DB 7; Length 41079;  
Best Local Similarity 90.0%; Pred. No. 2.2e+03;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
DB 2956 CATGTAATGCAGAAACATG 2975



```
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 653327
; LENGTH: 458
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-653327

Query Match      82.0%; Score 16.4; DB 4; Length 458;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
        ||||| ||||| ||||| |||||
Db      439 AGTAAATGCAGAAACAGG 456

RESULT 49
US-09-925-065A-222603/c
; Sequence 222603, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 222603
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-222603

Query Match      82.0%; Score 16.4; DB 4; Length 464;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
        ||||| ||||| ||||| |||||
Db      20 AGTAAATGCAGAAACAGG 3

RESULT 50
US-09-925-065A-150439/c
; Sequence 150439, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150440
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-150440

Query Match      82.0%; Score 16.4; DB 4; Length 503;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAG 19
        ||||| ||||| ||||| |||||
Db      42 AAGTAAATGCAGAAACAG 25

RESULT 51
US-09-925-065A-150440/c
; Sequence 150440, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single
; FILE REFERENCE: 108827.135
; CURRENT APPLICATION NUMBER: US/09/925,065A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: US 60/243,096
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/252,147
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,092
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/261,766
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/289,846
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 957086
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150440
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-065A-150440

Query Match      82.0%; Score 16.4; DB 4; Length 503;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 AAGTAAATGCAGAAACAG 19
        ||||| ||||| ||||| |||||
Db      42 AAGTAAATGCAGAAACAG 25

RESULT 52
US-10-027-632-6903
; Sequence 6903, Application US/10027632
```

```
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6903
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-6903

Query Match      82.0%; Score 16.4; DB 5; Length 504;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
Db      249 AGTAAATGCAGAAACAGG 266

RESULT 53
US-10-027-632-324576
; Sequence 324576, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 324576
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-324576

Query Match      82.0%; Score 16.4; DB 5; Length 504;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
Db      249 AGTAAATGCAGAAACAGG 266

RESULT 54
US-10-027-632-6903
; Sequence 6903, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6903
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-6903

Query Match      82.0%; Score 16.4; DB 6; Length 504;
Best Local Similarity 94.4%; Pred. No. 1.7e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 AGTAAATGCAGAAACAGG 20
Db      249 AGTAAATGCAGAAACAGG 266

RESULT 55
US-10-027-632-324576
; Sequence 324576, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
```

; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 324576  
; LENGTH: 504  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-324576

Query Match 82.0%; Score 16.4; DB 6; Length 504;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAGG 20  
|||||  
Db 249 AGTAAATGCAGAAACAGG 266

RESULT 56  
US-10-472-928-4807/C  
; Sequence 4807, Application US/10472928  
; Publication No. US20050020813A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH  
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE: P026926WO  
; CURRENT APPLICATION NUMBER: US/10/472,928  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: GB-0107658.7  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 4979  
; SOFTWARE: Seqwin99, version 1.03  
; SEQ ID NO 4807  
; LENGTH: 507  
; TYPE: DNA  
; ORGANISM: Streptococcus pneumoniae  
US-10-472-928-4807

Query Match 82.0%; Score 16.4; DB 8; Length 507;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 501 AAGTAAATGCAGTAAACAG 484

RESULT 57  
US-10-029-386-8721  
; Sequence 8721, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Ranzel, David K.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: ABOMICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 8721  
; LENGTH: 512  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC01311.11 SIGNAL = 1.2  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5

; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.73  
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4  
; OTHER INFORMATION: SWISSPROT HIT: P80665, EVALUE 4.80e+00  
; OTHER INFORMATION: EST HUMAN HIT: A1201312.1, EVALUE 8.00e-07  
; OTHER INFORMATION: NT HIT: g14729768, EVALUE 2.00e-06  
US-10-029-386-8721

Query Match 82.0%; Score 16.4; DB 6; Length 512;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 331 AAGCAATGCAGAAACAG 348

RESULT 58  
US-10-027-632-50588  
; Sequence 50588, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 50588  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-50588

Query Match 82.0%; Score 16.4; DB 5; Length 545;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 438 AAGTAAATGCAGAAACAG 455

RESULT 59  
US-10-027-632-50589  
; Sequence 50589, Application US/10027632  
; Publication No. US20020198371A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; TITLE OF INVENTION: Polymorphisms in the Human Genome  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20

; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 50588  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-50589

Query Match 82.0%; Score 16.4; DB 5; Length 545;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 438 AAGTAAATGCAGAAATCAG 455

RESULT 60  
US-10-027-632-50588  
; Sequence 50588, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108927.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 50588  
; LENGTH: 545  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-50588

Query Match 82.0%; Score 16.4; DB 6; Length 545;  
Best Local Similarity 94.4%; Pred. No. 1.7e+03;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 19  
|||||  
Db 438 AAGTAAATGCAGAAATCAG 455

Search completed: January 28, 2006, 01:45:44  
Job time : 350.966 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2006, 22:15:53 ; Search time 201.573 Seconds  
(without alignments)  
82.419 Million cell updates/sec

Title: US-10-716-005-4

Perfect score: 20

Sequence: 1 caagtaaatgcagaaacagg 20

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 6059551 seqs, 415333918 residues

Total number of hits satisfying chosen parameters: 12119102

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 60 summaries

Database :

- Published Applications NA New.\*
- 1: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
  - 2: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
  - 3: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
  - 4: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
  - 5: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
  - 6: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
  - 7: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
  - 8: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
  - 9: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
  - 10: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq3.\*
  - 11: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16.8	84.0	2866	7	US-10-750-185-39068
2	16.8	84.0	2866	7	US-10-750-623-39068
3	16.4	82.0	4320	7	US-10-750-185-27669
4	16.4	82.0	4320	7	US-10-750-623-27669
5	16.4	82.0	73072	8	US-11-124-368A-2919
6	16.4	82.0	149382	7	US-10-995-561-13272
7	15.8	79.0	201	7	US-10-995-561-43569
8	15.8	79.0	604	7	US-10-980-388-44
9	15.8	79.0	954	7	US-10-750-185-46466
10	15.8	79.0	954	7	US-10-750-185-46466
11	15.8	79.0	1108	7	US-10-750-185-45741
12	15.8	79.0	1108	7	US-10-750-623-45741
13	15.8	79.0	1223	7	US-10-750-185-56170
14	15.8	79.0	1223	7	US-10-750-623-56170
15	15.8	79.0	1435	7	US-10-750-185-52569
16	15.8	79.0	1435	7	US-10-750-623-52569
17	15.8	79.0	1447	7	US-10-750-185-61366
18	15.8	79.0	1447	7	US-10-750-623-61366
19	15.8	79.0	1770	8	US-11-128-061-572
20	15.8	79.0	1770	8	US-11-128-049-572
21	15.8	79.0	2659	7	US-10-750-185-29929
22	15.8	79.0	2659	7	US-10-750-623-29929

C	23	15.8	79.0	3001	8	US-11-145-703-194	Sequence 194, App
	24	15.8	79.0	21623	7	US-10-995-561-13497	Sequence 13497, A
C	25	15.8	79.0	44362	7	US-10-995-561-13422	Sequence 13422, A
C	26	15.8	79.0	119160	8	US-11-121-086-12	Sequence 12, Appl
C	27	15.8	79.0	156544	8	US-11-121-086-81	Sequence 81, Appl
C	28	15.8	79.0	159695	8	US-11-121-086-56	Sequence 56, Appl
C	29	15.8	79.0	394468	7	US-10-995-561-13473	Sequence 13473, A
C	30	15.8	79.0	1125000	7	US-10-995-561-13286	Sequence 13286, A
	31	15.4	77.0	598	7	US-10-750-185-19994	Sequence 19994, A
	32	15.4	77.0	598	7	US-10-750-623-19994	Sequence 19994, A
C	33	15.4	77.0	881	7	US-10-750-185-47634	Sequence 47634, A
C	34	15.4	77.0	881	7	US-10-750-623-47634	Sequence 47634, A
C	35	15.4	77.0	884	7	US-10-750-185-62704	Sequence 62704, A
C	36	15.4	77.0	884	7	US-10-750-623-62704	Sequence 62704, A
	37	15.4	77.0	1401	7	US-10-750-185-56765	Sequence 56765, A
	38	15.4	77.0	1401	7	US-10-750-623-56765	Sequence 56765, A
	39	15.4	77.0	1423	7	US-10-750-185-41820	Sequence 41820, A
	40	15.4	77.0	1423	7	US-10-750-623-41820	Sequence 41820, A
	41	15.4	77.0	2343	7	US-10-750-185-42759	Sequence 42759, A
	42	15.4	77.0	2343	7	US-10-750-623-42759	Sequence 42759, A
	43	15.4	77.0	2500	7	US-10-750-185-36275	Sequence 36275, A
	44	15.4	77.0	2500	7	US-10-750-623-36275	Sequence 36275, A
	45	15.4	77.0	185393	8	US-11-121-086-101	Sequence 101, Appl
	46	15.4	77.0	319608	8	US-11-145-703-1	Sequence 1, Appl
C	47	15.4	77.0	319608	8	US-11-145-703-1	Sequence 1, Appl
	48	15.4	77.0	1082144	8	US-11-117-187-211	Sequence 211, App
	49	15.2	76.0	25	8	US-11-136-527-326267	Sequence 326267, A
C	50	15.2	76.0	201	7	US-10-995-561-63705	Sequence 63705, A
	51	15.2	76.0	201	8	US-11-124-368A-16091	Sequence 16091, A
	52	15.2	76.0	201	8	US-11-124-368A-16143	Sequence 16143, A
	53	15.2	76.0	579	8	US-11-128-061-1990	Sequence 1990, Ap
	54	15.2	76.0	579	8	US-11-128-061-5632	Sequence 5632, Ap
	55	15.2	76.0	579	8	US-11-128-049-1990	Sequence 1990, Ap
	56	15.2	76.0	579	8	US-11-128-049-5632	Sequence 5632, Ap
	57	15.2	76.0	600	8	US-11-136-527-6925	Sequence 6925, Ap
	58	15.2	76.0	1115	7	US-10-750-185-28011	Sequence 28011, A
	59	15.2	76.0	1115	7	US-10-750-623-28011	Sequence 28011, A
C	60	15.2	76.0	1538	7	US-10-750-185-42218	Sequence 42218, A

ALIGNMENTS

RESULT 1  
US-10-750-185-39068  
; Sequence 39068, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 39068  
; LENGTH: 2866  
; TYPE: DNA  
; ORGANISM: Bovine 19866880837502  
US-10-750-185-39068

Query Match 84.0%; Score 16.8; DB 7; Length 2866;  
Best Local Similarity 90.0%; Pred. No. 1.se+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20  
|||||  
Db 235 CAAGTAAATGCAGCCACAGG 254

RESULT 2  
US-10-750-623-39068  
; Sequence 39068, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 39068  
; LENGTH: 2866  
; TYPE: DNA  
; ORGANISM: Bovine 19866881019429  
US-10-750-623-39068

Query Match 84.0%; Score 16.8; DB 7; Length 2866;  
Best Local Similarity 90.0%; Pred. No. 1.5e+02;  
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 20  
|||||  
Db 235 CAAGTAAATGCAGCCACAGG 254

RESULT 3  
US-10-750-185-27669/c  
; Sequence 27669, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27669  
; LENGTH: 4320  
; TYPE: DNA  
; ORGANISM: Bovine 19866881019429  
US-10-750-185-27669

Query Match 82.0%; Score 16.4; DB 7; Length 4320;  
Best Local Similarity 94.4%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 18  
|||||  
Db 4276 CAAGTAAATGCAGTAACA 4259

RESULT 4  
US-10-750-623-27669/c  
; Sequence 27669, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 27669  
; LENGTH: 4320  
; TYPE: DNA  
; ORGANISM: Bovine 19866881019429  
US-10-750-623-27669

Query Match 82.0%; Score 16.4; DB 7; Length 4320;  
Best Local Similarity 94.4%; Pred. No. 2.3e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAACAGG 18  
|||||  
Db 4276 CAAGTAAATGCAGTAACA 4259

RESULT 5  
US-11-124-368A-2919  
; Sequence 2919, Application US/11124368A  
; Publication No. US20050287559A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: James J. Devlin  
; APPLICANT: May Luke  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001524  
; CURRENT APPLICATION NUMBER: US/11/124,368A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,845  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/625,936  
; PRIOR FILING DATE: 2004-11-09  
; NUMBER OF SEQ ID NOS: 21112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2919  
; LENGTH: 73072  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-11-124-368A-2919

Query Match 82.0%; Score 16.4; DB 8; Length 73072;  
Best Local Similarity 94.4%; Pred. No. 3.2e+02;  
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAACAGG 20  
|||||  
Db 38284 AGAAAATGCAGAACAGG 38301

RESULT 6  
US-10-995-561-13272

```
; Sequence 13272, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13272
; LENGTH: 149382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(149382)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13272

Query Match      82.0%; Score 16.4; DB 7; Length 149382;
Best Local Similarity 94.4%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 19
Db      94792 AAGTAAATGCAGAAACAG 94809
|||||
RESULT 7
US-10-995-561-43569/c
; Sequence 43569, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43569
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-43569

Query Match      79.0%; Score 15.8; DB 7; Length 201;
Best Local Similarity 89.5%; Pred. No. 3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAAACAG 19
Db      197 CGAATTAATGCAGAAACAG 179
|||||
RESULT 8
US-10-980-388-44/c
; Sequence 44, Application US/10980388
; Publication No. US20050255490A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; APPLICANT: Parodi, Luis A.
; APPLICANT: Hiebsch, Ronald R.
; APPLICANT: Lind, Peter
; APPLICANT: Kaytes, Paul S.
; APPLICANT: Ruff, Valerie
; APPLICANT: Huff, Rica M.
; APPLICANT: Wood, Linda S.
```

```
; TITLE OF INVENTION: Novel G Protein-Coupled Receptors Cross-Reference To Related Appl
; FILE REFERENCE: 00325.US1
; CURRENT APPLICATION NUMBER: US/10/980,388
; CURRENT FILING DATE: 2004-11-02
; PRIOR APPLICATION NUMBER: US/09/791,932
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/184,305
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,304
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,303
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,397
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/184,247
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/188,880
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/217,369
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/217,370
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/218,492
; PRIOR FILING DATE: 2000-07-20
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 44
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-980-388-44

Query Match      79.0%; Score 15.8; DB 7; Length 604;
Best Local Similarity 89.5%; Pred. No. 3.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAAACAG 19
Db      94 CAAGTAAATGCAGAAACAG 76
|||||
RESULT 9
US-10-750-185-46466
; Sequence 46466, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MW1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46466
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Bovine 19866880562987
US-10-750-185-46466

Query Match      79.0%; Score 15.8; DB 7; Length 954;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 AAGTAAATGCAGAAACAG 20
```

```
Db      551 AAGTAAATTCAGGAACAGG 569
|||||
RESULT 10
US-10-750-623-46466
; Sequence 46466, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 46466
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Bovine 19866880562987
US-10-750-623-46466

Query Match      79.0%; Score 15.8; DB 7; Length 954;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 AAGTAAATTCAGGAACAGG 20
|||||
Db      551 AAGTAAATTCAGGAACAGG 569

RESULT 11
US-10-750-185-45741/c
; Sequence 45741, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 45741
; LENGTH: 1108
; TYPE: DNA
; ORGANISM: Bovine 19866880792232
US-10-750-185-45741

Query Match      79.0%; Score 15.8; DB 7; Length 1108;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAAACAG 19
|||||
Db      250 CAAGAAATGCAGACACAG 232

RESULT 12
US-10-750-623-45741/c
; Sequence 45741, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 45741
; LENGTH: 1108
; TYPE: DNA
; ORGANISM: Bovine 19866880792232
US-10-750-623-45741

Query Match      79.0%; Score 15.8; DB 7; Length 1108;
Best Local Similarity 89.5%; Pred. No. 3.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAAACAG 19
|||||
Db      250 CAAGAAATGCAGACACAG 232

RESULT 13
US-10-750-185-56170/c
; Sequence 56170, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 56170
; LENGTH: 1223
; TYPE: DNA
; ORGANISM: Bovine 19866880790313
US-10-750-185-56170

Query Match      79.0%; Score 15.8; DB 7; Length 1223;
Best Local Similarity 89.5%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 CAAGTAAATGCAGAAACAG 19
|||||
Db      289 CAATAAATGCAGATACAG 271

RESULT 14
US-10-750-623-56170/c
```

```
; Sequence 56170, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 56170
; LENGTH: 1223
; TYPE: DNA
; ORGANISM: Bovine 19866880961908
; US-10-750-623-56170

Query Match          79.0%; Score 15.8; DB 7; Length 1223;
Best Local Similarity 89.5%; Pred. No. 3.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   |||||
Db 289 CAAATAAATGCAGATACAG 271

RESULT 15
US-10-750-185-52569/c
; Sequence 52569, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52569
; LENGTH: 1435
; TYPE: DNA
; ORGANISM: Bovine 19866880961908
; US-10-750-185-52569

Query Match          79.0%; Score 15.8; DB 7; Length 1435;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 20
   |||||
Db 648 AAGTAAATGCAGAAATAAG 630

RESULT 16
US-10-750-623-52569/c
; Sequence 52569, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
```

```
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 52569
; LENGTH: 1435
; TYPE: DNA
; ORGANISM: Bovine 19866880961908
; US-10-750-623-52569

Query Match          79.0%; Score 15.8; DB 7; Length 1435;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 20
   |||||
Db 648 AAGTAAATGCAGAAATAAG 630

RESULT 17
US-10-750-185-61366/c
; Sequence 61366, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 61366
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Bovine 19866880352244
; US-10-750-185-61366

Query Match          79.0%; Score 15.8; DB 7; Length 1447;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
   |||||
Db 1008 CAGGAAATGCAGAAACAG 990

RESULT 18
US-10-750-623-61366/c
; Sequence 61366, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
```

```
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61366
; LENGTH: 1447
; TYPE: DNA
; ORGANISM: Bovine 19866880352244
US-10-750-623-61366

Query Match          79.0%; Score 15.8; DB 7; Length 1447;
Best Local Similarity 89.5%; Pred. No. 3.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
   ||| ||||| ||||| |||||
Db 1008 CAGGAAATGCAGAAACAG 990

RESULT 19
US-11-128-061-472
; Sequence 572, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 572
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Mesocricetus auratus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1058)..(1088)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-061-572

Query Match          79.0%; Score 15.8; DB 8; Length 1770;
Best Local Similarity 89.5%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 20
   ||| ||||| ||||| |||||
Db 839 AAATAAATGCAGAAACAG 857

RESULT 20
US-11-128-049-572
; Sequence 572, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 572
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Mesocricetus auratus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1058)..(1088)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-572

Query Match          79.0%; Score 15.8; DB 8; Length 1770;
Best Local Similarity 89.5%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 20
   ||| ||||| ||||| |||||
Db 839 AAATAAATGCAGAAACAG 857

RESULT 20
US-11-128-049-572
; Sequence 572, Application US/11128049
; Publication No. US20060010513A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR
; FILE REFERENCE: 01997.027700
; CURRENT APPLICATION NUMBER: US/11/128,049
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
; PRIOR FILING DATE: 2004-05-11
; NUMBER OF SEQ ID NOS: 7285
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 572
; LENGTH: 1770
; TYPE: DNA
; ORGANISM: Mesocricetus auratus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1058)..(1088)
; OTHER INFORMATION: n is a, c, g, or t
US-11-128-049-572

Query Match          79.0%; Score 15.8; DB 8; Length 1770;
Best Local Similarity 89.5%; Pred. No. 3.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 AAGTAAATGCAGAAACAG 20
   ||| ||||| ||||| |||||
Db 839 AAATAAATGCAGAAACAG 857

RESULT 22
US-10-750-623-29929
; Sequence 29929, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM11100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29929
; LENGTH: 2659
; TYPE: DNA
; ORGANISM: Bovine 19866880628714
US-10-750-185-29929

Query Match          79.0%; Score 15.8; DB 7; Length 2659;
Best Local Similarity 89.5%; Pred. No. 4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 CAAGTAAATGCAGAAACAG 19
   ||||| ||||| ||||| |||||
Db 601 CAAGTAGATGGAGAAACAG 619

RESULT 22
US-10-750-623-29929
```

```
; Sequence 29929, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 29929
; LENGTH: 2659
; TYPE: DNA
; ORGANISM: Bovine 1986680628714
US-10-750-623-29929

Query Match          79.0%; Score 15.8; DB 7; Length 2659;
Best Local Similarity 89.5%; Pred. No. 4e+02; 0; Mismatches 0; Indels 0; Gaps 0;
Matches 17; Conservative 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 601 CAAGTAGATGCAGAAACAG 619

RESULT 23
US-11-145-703-194/c
; Sequence 194, Application US/11145703
; Publication No. US20050260667A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53.US16.DIV
; CURRENT APPLICATION NUMBER: US/11/145,703
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US/10/147,603
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 234
; SOFTWARE: Patent.pn
; SEQ ID NO 194
; LENGTH: 3001
```

```
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1501
; OTHER INFORMATION: 99-26223-225 : polymorphic base G or T
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1481..1500
; OTHER INFORMATION: 99-26223-225.mis1,
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1502..1520
; OTHER INFORMATION: 99-26223-225.mis2, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1277..1297
; OTHER INFORMATION: upstream amplification primer
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1842..1862
; OTHER INFORMATION: downstream amplification primer, complement
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1489..1513
; OTHER INFORMATION: 99-26223-225 probe
US-11-145-703-194

Query Match          79.0%; Score 15.8; DB 8; Length 3001;
Best Local Similarity 89.5%; Pred. No. 4.1e+02; 0; Mismatches 2; Indels 0; Gaps 0;
Matches 17; Conservative 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 1917 CAATTAAATGCAGAAACAG 1899

RESULT 24
US-10-995-561-13497
; Sequence 13497, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13497
; LENGTH: 21623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-13497

Query Match          79.0%; Score 15.8; DB 7; Length 21623;
Best Local Similarity 89.5%; Pred. No. 5.1e+02; 0; Mismatches 2; Indels 0; Gaps 0;
Matches 17; Conservative 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||||| ||||| ||||| |||||
Db 17255 CAATTAAATGCAGAAACAG 17273

RESULT 25
US-10-995-561-13422/c
; Sequence 13422, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
```

```
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13422
; LENGTH: 44362
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(44362)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13422

Query Match          79.0%; Score 15.8; DB 7; Length 44362;
Best Local Similarity 89.5%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||| ||||| ||||| |||||
Db 11504 CAATTAATGCAGAAACAG 11486

RESULT 26
US-11-121-086-12/c
; Sequence 12, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 12
; LENGTH: 119160
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-12

Query Match          79.0%; Score 15.8; DB 8; Length 119160;
Best Local Similarity 89.5%; Pred. No. 6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||| ||||| ||||| |||||
Db 80153 CAATTAATGCAGAAACAG 80135

RESULT 27
US-11-121-086-81/c
; Sequence 81, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 81
; LENGTH: 156544
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-81

Query Match          79.0%; Score 15.8; DB 8; Length 156544;
Best Local Similarity 89.5%; Pred. No. 6.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||| ||||| ||||| |||||
Db 68402 CAATCAATGCAGAAACAG 68384

RESULT 28
US-11-121-086-56/c
; Sequence 56, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 56
; LENGTH: 159695
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-56

Query Match          79.0%; Score 15.8; DB 8; Length 159695;
Best Local Similarity 89.5%; Pred. No. 6.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19
    ||| ||||| ||||| |||||
Db 80674 CAATTAATGCAGAAACAG 80656

RESULT 29
US-10-995-561-13473/c
; Sequence 13473, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13473
; LENGTH: 394468
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(394468)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-995-561-13473

Query Match          79.0%; Score 15.8; DB 7; Length 394468;
Best Local Similarity 89.5%; Pred. No. 6.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACAG 20
    ||||| ||||| ||||| |||||
```



Db 171081 AGTAAATGCTAAACAGG 171063

RESULT 30  
US-10-995-561-13286/c  
; Sequence 13286, Application US/10995561  
; Publication No. US2005072054A1  
; GENERAL INFORMATION:  
; APPLICANT: CARGILL, Michele et al.  
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF  
; TITLE OF INVENTION: DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001559  
; CURRENT APPLICATION NUMBER: US/10/995,561  
; CURRENT FILING DATE: 2004-11-24  
; NUMBER OF SEQ ID NOS: 85702  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13286  
; LENGTH: 1125000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(1125000)  
; OTHER INFORMATION: n is any nucleotide  
US-10-995-561-13286

Query Match 79.0%; Score 15.8; DB 7; Length 1125000;  
Best Local Similarity 89.5%; Pred. No. 5e+02;  
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAG 19  
Db 657182 CGAATAATGCAGAAACAG 657164

RESULT 31  
US-10-750-185-19994  
; Sequence 19994, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 19994  
; LENGTH: 598  
; TYPE: DNA  
; ORGANISM: Bovine MMBT01307  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(101)  
; OTHER INFORMATION: n is any nucleotide  
US-10-750-185-19994

Query Match 77.0%; Score 15.4; DB 7; Length 598;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19  
Db 307 AGTAAATGCAGAAACAG 323

RESULT 32  
US-10-750-623-19994  
; Sequence 19994, Application US/10750623  
; Publication No. US20050287531A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-1  
; CURRENT APPLICATION NUMBER: US/10/750,623  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 19994  
; LENGTH: 598  
; TYPE: DNA  
; ORGANISM: Bovine MMBT01307  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(101)  
; OTHER INFORMATION: n is any nucleotide  
US-10-750-623-19994

Query Match 77.0%; Score 15.4; DB 7; Length 598;  
Best Local Similarity 94.1%; Pred. No. 5.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19  
Db 307 AGTAAATGCAGAAACAG 323

RESULT 33  
US-10-750-185-47634/c  
; Sequence 47634, Application US/10750185  
; Publication No. US20050260603A1  
; GENERAL INFORMATION:  
; APPLICANT: MMI GENOMICS, INC.  
; APPLICANT: DENISE, Sue K.  
; APPLICANT: KERR, Richard  
; APPLICANT: ROSENFELD, David  
; APPLICANT: HOLM, Tom  
; APPLICANT: BATES, Stephen  
; APPLICANT: FANTIN, Dennis  
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS  
; FILE REFERENCE: MM1100-2  
; CURRENT APPLICATION NUMBER: US/10/750,185  
; CURRENT FILING DATE: 2003-12-31  
; PRIOR APPLICATION NUMBER: US 60/437,482  
; PRIOR FILING DATE: 2002-12-31  
; NUMBER OF SEQ ID NOS: 64922  
; SOFTWARE: PatentIN version 3.1  
; SEQ ID NO 47634  
; LENGTH: 881  
; TYPE: DNA  
; ORGANISM: Bovine 19866881682228  
US-10-750-185-47634

Query Match 77.0%; Score 15.4; DB 7; Length 881;  
Best Local Similarity 94.1%; Pred. No. 5.4e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19  
Db 307 AGTAAATGCAGAAACAG 323



```
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 56765
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Bovine 19866881596118
US-10-750-623-56765

Query Match 77.0%; Score 15.4; DB 7; Length 1401;
Best Local Similarity 94.1%; Pred. No. 5.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACA 18
Db 474 AAGTAAATGCAGAAACA 490

RESULT 39
US-10-750-185-41820
; Sequence 41820, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41820
; LENGTH: 1423
; TYPE: DNA
; ORGANISM: Bovine 19866881002825
US-10-750-185-41820

Query Match 77.0%; Score 15.4; DB 7; Length 1423;
Best Local Similarity 94.1%; Pred. No. 5.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACA 18
Db 1108 AAGTAAATGCAGAAACA 1124

RESULT 40
US-10-750-623-41820
; Sequence 41820, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 41820
; LENGTH: 1423
; TYPE: DNA
; ORGANISM: Bovine 19866881002825
US-10-750-623-41820

Query Match 77.0%; Score 15.4; DB 7; Length 1423;
Best Local Similarity 94.1%; Pred. No. 5.7e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AAGTAAATGCAGAAACA 18
Db 1108 AAGTAAATGCAGAAACA 1124

RESULT 41
US-10-750-185-42759
; Sequence 42759, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 42759
; LENGTH: 2343
; TYPE: DNA
; ORGANISM: Bovine 19866880345005
US-10-750-185-42759

Query Match 77.0%; Score 15.4; DB 7; Length 2343;
Best Local Similarity 94.1%; Pred. No. 6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAAC 17
Db 1301 CAAGTAAATGCAGAAAC 1317

RESULT 42
US-10-750-623-42759
; Sequence 42759, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
```

```
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; PRIOR FILING DATE: 2003-12-31
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 36275
; LENGTH: 2500
; TYPE: DNA
; ORGANISM: Bovine 19866880753661
US-10-750-623-42759

Query Match          77.0%; Score 15.4; DB 7; Length 2343;
Best Local Similarity 94.1%; Pred. No. 6e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAAC 17
   ||||| ||||| |||||
Db 1301 CAAGTAAAGCAGAAAC 1317

RESULT 43
US-10-750-185-46275
; Sequence 36275, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 36275
; LENGTH: 2500
; TYPE: DNA
; ORGANISM: Bovine 19866880753661
US-10-750-185-36275

Query Match          77.0%; Score 15.4; DB 7; Length 2500;
Best Local Similarity 94.1%; Pred. No. 6.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 206 AGTAAATGCAGAAACAG 222

RESULT 44
US-10-750-623-36275
; Sequence 36275, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
```

```
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIN version 3.1
; SEQ ID NO 36275
; LENGTH: 2500
; TYPE: DNA
; ORGANISM: Bovine 19866880753661
US-10-750-623-36275

Query Match          77.0%; Score 15.4; DB 7; Length 2500;
Best Local Similarity 94.1%; Pred. No. 6.1e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 206 AGTAAATGCAGAAACAG 222

RESULT 45
US-11-121-086-101
; Sequence 101, Application US/11121086
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138.6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; CURRENT FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIN version 3.3
; SEQ ID NO 101
; LENGTH: 185393
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-121-086-101

Query Match          77.0%; Score 15.4; DB 8; Length 185393;
Best Local Similarity 94.1%; Pred. No. 9e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 AGTAAATGCAGAAACAG 19
   ||||| ||||| |||||
Db 12409 AGTAAATGTAGAAACAG 12425

RESULT 46
US-11-145-703-1
; Sequence 1, Application US/11145703
; Publication No. US20050260667A1
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53.US16.DIV
; CURRENT APPLICATION NUMBER: US/11/145,703
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US/10/147,603
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
```

```

; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 234
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 65854..67854
; OTHER INFORMATION: 3'regulatory region g35018 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 214676..214793

```

```

; OTHER INFORMATION: exon T g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216836..216915
; OTHER INFORMATION: exon V g35030 gene
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 213818..215818
; OTHER INFORMATION: 3'regulatory region g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215941
; OTHER INFORMATION: exon R complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215819..215975
; OTHER INFORMATION: exon Rbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..216952
; OTHER INFORMATION: exon Qbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 216661..217061
; OTHER INFORMATION: exon Q complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 217027..217061
; OTHER INFORMATION: exon Q1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 229647..229742
; OTHER INFORMATION: exon X complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 230408..230721
; OTHER INFORMATION: exon P complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231272..231412
; OTHER INFORMATION: exon Obis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231787..231880
; OTHER INFORMATION: exon O2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 231870..231879
; OTHER INFORMATION: exon O1 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 234174..234321
; OTHER INFORMATION: exon O complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 237406..237428
; OTHER INFORMATION: exon Nbis complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239807
; OTHER INFORMATION: exon N2 complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 239719..239853
; OTHER INFORMATION: exon N complement g34872 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240569
; OTHER INFORMATION: exon M1117 complement g34872 gene

```

```
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240596
; OTHER INFORMATION: exon M1090 complement g34872 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240617
; OTHER INFORMATION: exon M1069 complement g34872 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240644
; OTHER INFORMATION: exon M52 complement g34872 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240824
; OTHER INFORMATION: exon M862 complement g34872 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240528..240994
; OTHER INFORMATION: exon M692 complement g34872 gene
;
; FEATURE:
; NAME/KEY: 4xon
; LOCATION: 240528..241685
; OTHER INFORMATION: exon M1 complement g34872 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 240800..240993
; OTHER INFORMATION: exon M51 complement g34872 gene
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 241686..243685
; OTHER INFORMATION: 5'regulatory region g34872 gene
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 290652..292652
; OTHER INFORMATION: 3'regulatory region g34665 gene
;
;
; Query Match 77.0%; Score 15.4; DB 8; Length 319608;
; Best Local Similarity 94.1%; Pred. No. 8.e+02;
; Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; Qy 2 AAGTAAATGCAGAAACA 18
;
; Db 161192 AAGTGAATGCAGAAACA 161208
;
;
; RESULT 47
; US-11-145-703-1/c
; Sequence 1, Application US/11145703
; Publication No. US20050260667A1
;
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilva
; APPLICANT: Bouqueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53.US16.DIV
; CURRENT APPLICATION NUMBER: US/11/145,703
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US/10/147,603
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
```

```
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 234
; SOFTWARE: Patent.Pm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 31..1107
; OTHER INFORMATION: 5'regulatory region g35018 gene
;
; NAME/KEY: exon
; LOCATION: 1108..1289
; OTHER INFORMATION: exon A g35018 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 14877..14920
; OTHER INFORMATION: exon B g35018 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 18778..18862
; OTHER INFORMATION: exon Bbis g35018 gene
;
; NAME/KEY: exon
; LOCATION: 25593..25740
; OTHER INFORMATION: exon C g35018 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 29388..29502
; OTHER INFORMATION: exon D g35018 gene
;
; NAME/KEY: exon
; LOCATION: 29967..30282
; OTHER INFORMATION: exon E g35018 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 64666..64812
; OTHER INFORMATION: exon F g35018 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 65505..65853
; OTHER INFORMATION: exon G g35018 gene
;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 65854..67854
; OTHER INFORMATION: 3'regulatory region g35018 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 94124..94964
; OTHER INFORMATION: exon g35017
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 201188..201234
; OTHER INFORMATION: exon S g35030 gene
;
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T g35030 gene
;
; FEATURE:
; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U g35030 gene
;
; NAME/KEY: exon
```

LOCATION: 216836..216915  
OTHER INFORMATION: exon V g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 213818..215818  
OTHER INFORMATION: 3'regulatory region g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215819..215941  
OTHER INFORMATION: exon R complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215819..215975  
OTHER INFORMATION: exon Rbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216661..216952  
OTHER INFORMATION: exon Qbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216661..217061  
OTHER INFORMATION: exon Q complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 217027..217061  
OTHER INFORMATION: exon Q1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 229647..229742  
OTHER INFORMATION: exon X complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 230408..230721  
OTHER INFORMATION: exon P complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231272..231412  
OTHER INFORMATION: exon Obis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231787..231880  
OTHER INFORMATION: exon O2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231870..231879  
OTHER INFORMATION: exon O1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 234174..234321  
OTHER INFORMATION: exon O complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 237406..237428  
OTHER INFORMATION: exon Nbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239807  
OTHER INFORMATION: exon N2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239853  
OTHER INFORMATION: exon N complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240569  
OTHER INFORMATION: exon M1117 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240596  
OTHER INFORMATION: exon M1090 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240617

OTHER INFORMATION: exon M1069 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240644  
OTHER INFORMATION: exon MS2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240824  
OTHER INFORMATION: exon M862 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240994  
OTHER INFORMATION: exon M692 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..241685  
OTHER INFORMATION: exon M1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240800..240993  
OTHER INFORMATION: exon MS1 complement g34872 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 241686..243885  
OTHER INFORMATION: 5'regulatory region g34872 gene  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 290652..292652  
OTHER INFORMATION: 3'regulatory region g34665 gene  
Query Match 77.0%; Score 15.4; DB 8; Length 319608;  
Best Local Similarity 94.1%; Pred. No. 8.8e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 4 GTAAATGCAGAAACAGG 20  
Db 193336 GTAAATGTAGAAACAGG 193320  
RESULT 48  
US-11-117-187-211  
; Sequence 211, Application US/11117187  
; Publication No. US20050266560A1  
; GENERAL INFORMATION:  
; APPLICANT: PREUSS, DAPHNE  
; APPLICANT: COPENHAVER, GREGORY  
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS  
; FILE REFERENCE: ARCD:309US  
; CURRENT APPLICATION NUMBER: US/11/117,187  
; CURRENT FILING DATE: 2005-04-28  
; PRIOR APPLICATION NUMBER: US/09/531,120  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/125,219  
; PRIOR FILING DATE: 1999-03-18  
; NUMBER OF SEQ ID NOS: 212  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 211  
; LENGTH: 1082144  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-11-117-187-211  
Query Match 77.0%; Score 15.4; DB 8; Length 1082144;  
Best Local Similarity 94.1%; Pred. No. 6.1e+02;  
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 CAAAGTAAATGCAGAAAC 17  
Db 190955 CAAAGTAAATACAGAAAC 190971  
RESULT 49  
US-11-136-527-326267

```
; Sequence 326267, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 326267
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-326267

Query Match 76.0%; Score 15.2; DB 8; Length 25;
Best Local Similarity 85.0%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 1 CAAGTTAATGCAGGAACCG 20

RESULT 50
US-10-995-561-63705/c
; Sequence 63705, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63705
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-63705

Query Match 76.0%; Score 15.2; DB 7; Length 201;
Best Local Similarity 85.0%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 161 CAAGTAAATGCAGAAACAGG 142

RESULT 51
US-11-124-368A-16091
; Sequence 16091, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845

; Sequence 326267, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 326267
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-326267

Query Match 76.0%; Score 15.2; DB 8; Length 25;
Best Local Similarity 85.0%; Pred. No. 4.3e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 1 CAAGTTAATGCAGGAACCG 20

RESULT 50
US-10-995-561-63705/c
; Sequence 63705, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63705
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-995-561-63705

Query Match 76.0%; Score 15.2; DB 7; Length 201;
Best Local Similarity 85.0%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 161 CAAGTAAATGCAGAAACAGG 142

RESULT 51
US-11-124-368A-16091
; Sequence 16091, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845

; Sequence 326267, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 326267
; LENGTH: 25
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Probe
US-11-136-527-326267

Query Match 76.0%; Score 15.2; DB 8; Length 201;
Best Local Similarity 85.0%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 71 CAAGGCAATGCTGAAACAGG 90

RESULT 52
US-11-124-368A-16143
; Sequence 16143, Application US/11124368A
; Publication No. US20050287559A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: James J. Devlin
; APPLICANT: May Luke
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Vascular Diseases, Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001524
; CURRENT APPLICATION NUMBER: US/11/124,368A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,845
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/625,936
; PRIOR FILING DATE: 2004-11-09
; NUMBER OF SEQ ID NOS: 21112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16143
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-124-368A-16143

Query Match 76.0%; Score 15.2; DB 8; Length 201;
Best Local Similarity 85.0%; Pred. No. 5.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
Db 121 CAAGGCAATGCTGAAACAGG 140

RESULT 53
US-11-128-061-1990
; Sequence 1990, Application US/11128061
; Publication No. US20060003958A1
; GENERAL INFORMATION:
; APPLICANT: Melville, Mark W.
; APPLICANT: Charlebois, Timothy S.
; APPLICANT: Mounts, William M.
; APPLICANT: Hann, Louane E.
; APPLICANT: Sinacore, Martin S.
; APPLICANT: Leonard, Mark W.
; APPLICANT: Brown, Eugene L.
; APPLICANT: Miller, Christopher P.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION
; FILE REFERENCE: 01997.027701
; CURRENT APPLICATION NUMBER: US/11/128,061
; CURRENT FILING DATE: 2005-05-11
; PRIOR APPLICATION NUMBER: US 60/570,425
```



;  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1990  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-061-1990

Query Match 76.0%; Score 15.2; DB 8; Length 579;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
DB 539 CAAGTCAATGCAGAAATCATG 558

RESULT 54

US-11-128-061-5632  
; Sequence 5632, Application US/11128061  
; Publication No. US20060003958A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES RELATED TO OLIGONUCLEOTIDE ARRAYS  
; TITLE OF INVENTION: TO MONITOR GENE EXPRESSION  
; FILE REFERENCE: 01997.027701  
; CURRENT APPLICATION NUMBER: US/11/128,061  
; CURRENT FILING DATE: 2005-05-11  
; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 5632  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-061-5632

Query Match 76.0%; Score 15.2; DB 8; Length 579;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
DB 539 CAAGTCAATGCAGAAATCATG 558

RESULT 55

US-11-128-049-1990  
; Sequence 1990, Application US/11128049  
; Publication No. US20060010513A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FOR  
; TITLE OF INVENTION: MAKING AND USING SAME  
; FILE REFERENCE: 01997.027700  
; CURRENT APPLICATION NUMBER: US/11/128,049  
; CURRENT FILING DATE: 2005-05-11

;  
; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 1990  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-049-1990

Query Match 76.0%; Score 15.2; DB 8; Length 579;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
DB 539 CAAGTCAATGCAGAAATCATG 558

RESULT 56

US-11-128-049-5632  
; Sequence 5632, Application US/11128049  
; Publication No. US20060010513A1  
; GENERAL INFORMATION:  
; APPLICANT: Melville, Mark W.  
; APPLICANT: Charlebois, Timothy S.  
; APPLICANT: Mounts, William M.  
; APPLICANT: Hann, Louane E.  
; APPLICANT: Sinacore, Martin S.  
; APPLICANT: Leonard, Mark W.  
; APPLICANT: Brown, Eugene L.  
; APPLICANT: Miller, Christopher P.  
; TITLE OF INVENTION: OLIGONUCLEOTIDE ARRAYS TO MONITOR GENE EXPRESSION AND METHODS FO  
; TITLE OF INVENTION: MAKING AND USING SAME  
; FILE REFERENCE: 01997.027700  
; CURRENT APPLICATION NUMBER: US/11/128,049  
; CURRENT FILING DATE: 2005-05-11  
; PRIOR APPLICATION NUMBER: US 60/570,425  
; PRIOR FILING DATE: 2004-05-11  
; NUMBER OF SEQ ID NOS: 7285  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 5632  
; LENGTH: 579  
; TYPE: DNA  
; ORGANISM: Cricetulus griseus  
US-11-128-049-5632

Query Match 76.0%; Score 15.2; DB 8; Length 579;  
Best Local Similarity 85.0%; Pred. No. 6.3e+02;  
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20  
||||| ||||| ||||| ||||| |||||  
DB 539 CAAGTCAATGCAGAAATCATG 558

RESULT 57

US-11-136-527-6925  
; Sequence 6925, Application US/11136527  
; Publication No. US20050287570A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; APPLICANT: Mounts, William M.  
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes  
; FILE REFERENCE: 031896-041000 (AM101086)  
; CURRENT APPLICATION NUMBER: US/11/136,527  
; CURRENT FILING DATE: 2005-05-25  
; PRIOR APPLICATION NUMBER: US 60/574,294  
; PRIOR FILING DATE: 2005-05-26  
; NUMBER OF SEQ ID NOS: 362830  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 6925  
; LENGTH: 600

```

; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6925

Query Match
Best Local Similarity 76.0%; Score 15.2; DB 8; Length 600;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
Db 527 CAAGTTAATGCAGGAACCG 546

RESULT 58
US-10-750-185-28011
; Sequence 28011, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28011
; LENGTH: 1115
; TYPE: DNA
; ORGANISM: Bovine 1986680751726
US-10-750-185-28011

Query Match
Best Local Similarity 76.0%; Score 15.2; DB 7; Length 1115;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
Db 1083 CAAATAAATGGGAAACAGG 1102

RESULT 59
US-10-750-623-28011
; Sequence 28011, Application US/10750623
; Publication No. US20050287531A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: METHODS AND SYSTEMS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-1
; CURRENT APPLICATION NUMBER: US/10/750,623
; CURRENT FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28011
; LENGTH: 1115
; TYPE: DNA
; ORGANISM: Bovine 1986680751726
US-10-750-623-28011

; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-6925

Query Match
Best Local Similarity 76.0%; Score 15.2; DB 7; Length 1115;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
Db 1083 CAAATAAATGGGAAACAGG 1102

RESULT 60
US-10-750-185-42218/c
; Sequence 42218, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR APPLICATION NUMBER: US 60/437,482
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42218
; LENGTH: 1538
; TYPE: DNA
; ORGANISM: Bovine 19866808081480
US-10-750-185-42218

Query Match
Best Local Similarity 76.0%; Score 15.2; DB 7; Length 1538;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 CAAGTAAATGCAGAAACAGG 20
   ||||| ||||| ||||| |||||
Db 443 CAAGTCAAAACAGAAACAGG 424

Search completed: January 28, 2006, 02:00:54
Job time : 208.823 secs
```